

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification:

A61K

A2

(11) International Publication Number:

WO 99/42076

(43) International Publication Date:

20 August 1999 (26.08.99)

(21) International Application Number:

PCT/US99/03268

(22) International Filing Date:

17 February 1999 (17.02.99)

(30) Priority Data:

09/025,197 18 February 1998 (18.02.98) US
09/072,967 5 May 1998 (05.05.98) US

(71) Applicant: CORINA CORPORATION [US/US], 1124 Columbia Street, Seattle, WA 98104 (US).

(72) Inventors: REED, Steven, G., 2845 122nd Place, N.E., Bellevue, WA 98005 (US); SKEIKY, Yasin, A., W., 8327 25th Street, Seattle, WA 98107 (US); DILLON, David, C., 21607 N.E. 24th Street, Redmond, WA 98053 (US); CAMPOS-NEIRO, Antonio, 9308 Midship Court, N.E., Bainbridge Island, WA 98110 (US); HOUGHTON, Raymond, 2636 242nd Place, S.E., Bothell, WA 98021 (US); VEDVICK, Thomas, S., 124 South 300th Place, Federal Way, WA 98003 (US); TWARDZIK, Daniel, R., 10195 South Beach Drive, Bainbridge Island, WA 98110 (US); LODES, Michael, J., 9223 36th Avenue S.W., Seattle, WA 98126 (US); HENDRICKSON, Ronald, C., 4114 S.W. Charlestown Street, Seattle, WA 98116 (US).

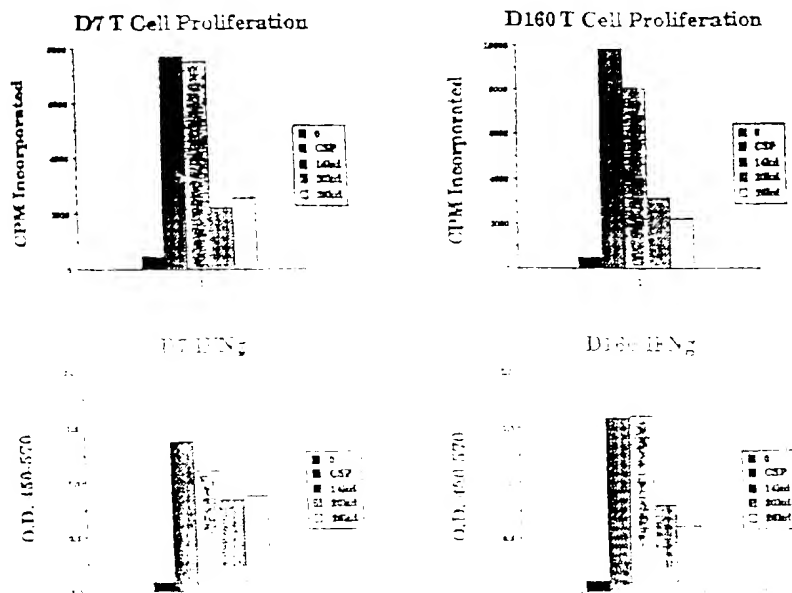
(74) Agents: CORUZZI, Laura, A. et al., Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IL, IT, MC, NL, PT, SE), OAPI patent (BF, BI, CE, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report

(54) Title: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCI on the front pages of pamphlets publishing international applications under the PCI

AI	Angola	IS	Israel	IS	Israel	SI	Slovenia
AM	Armenia	IT	Italy	LI	Lithuania	SK	Slovakia
AT	Austria	FR	France	LT	Luxembourg	SN	Senegal
AX	Aland Islands	GA	Gabon	LV	Latvia	SZ	Switzerland
AZ	Azerbaijan	GB	Great Britain	MG	Madagascar	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Moldova	TD	Chad
BB	Barbados	GH	Ghana	MG	Madagascar	TL	Timor-Leste
BE	Belgium	GN	Guinea	MK	Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HR	Croatia	ML	Mali	TT	Trinidad and Tobago
BI	Burundi	IE	Ireland	MN	Mongolia	UA	Ukraine
BJ	Benin	IL	Israel	MR	Morocco	UG	Uganda
BY	Belarus	IS	Israel	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MY	Malaysia	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NI	Nicaragua	VA	Vatican
CG	Congo	KE	Kenya	NI	Nicaragua	YC	Yemen
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Ivory Coast	KP	Korea, North	NZ	New Zealand		
CM	Cameroun			P	Poland		

COMPOUNDS AND METHODS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS

5 CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of U.S. Application No. 9/025,197, filed February 18, 1998; which is a continuation-in-part of U.S. Application No. 08/942,578, filed October 1, 1997; which is a continuation-in-part of U.S. Application No. 08/818,112, filed March 13, 1997; which is a continuation-in-part of U.S. Application No. 08/730,510, filed October 11, 1996; which claims priority from PCT Application No. PCT/US 96/14674, filed August 30, 1996, and is a continuation-in-part of U.S. Application No. 08/680,574, filed July 12, 1996; which is a continuation-in-part of U.S. Application No. 08/659,683, filed June 5, 1996; which is a continuation-in-part of U.S. Application No. 08/620,874, filed March 22, 1996, now abandoned; which is a continuation-in-part of U.S. Application No. 08/533,634, filed September 22, 1995, now abandoned; which is a continuation-in-part of U.S. Application No. 08/523,436, filed September 1, 1995, now abandoned.

TECHNICAL FIELD

20 The present invention relates generally to detecting, treating and preventing *Mycobacterium tuberculosis* infection. The invention is more particularly related to polypeptides comprising a *Mycobacterium tuberculosis* antigen, or a portion or other variant thereof, and the use of such polypeptide for diagnosing and vaccinating against *Mycobacterium tuberculosis* infection.

BACKGROUND OF THE INVENTION

25 Tuberculosis is a chronic infectious disease that is generally caused by infection with *Mycobacterium tuberculosis*. It is a major disease in developing countries as well as in industrialized countries. Tuberculosis is caused by

manifested as an acute inflammation of the lungs, resulting in fever and a nonproductive cough. If left untreated, serious complications and death typically result.

Although tuberculosis can generally be controlled using extended antibiotic therapy, such treatment is not sufficient to prevent the spread of the disease. Infected individuals may be asymptomatic, but contagious, for some time. In addition, although compliance with the treatment regimen is critical, patient behavior is difficult to monitor. Some patients do not complete the course of treatment, which can lead to ineffective treatment and the development of drug resistance.

Inhibiting the spread of tuberculosis requires effective vaccination and accurate, early diagnosis of the disease. Currently, vaccination with live bacteria is the most efficient method for inducing protective immunity. The most common *Mycobacterium* employed for this purpose is *Bacillus Calmette-Guerin* (BCG), an avirulent strain of *Mycobacterium bovis*. However, the safety and efficacy of BCG is a source of controversy and some countries, such as the United States, do not vaccinate the general public. Diagnosis is commonly achieved using a skin test, which involves intradermal exposure to tuberculin PPD (protein-purified derivative). Antigen-specific T cell responses result in measurable induration at the injection site by 48-72 hours after injection, which indicates exposure to *Mycobacterial* antigens. Sensitivity and specificity have, however, been a problem with this test, and individuals vaccinated with BCG cannot be distinguished from infected individuals.

While macrophages have been shown to act as the principal effectors of *M. tuberculosis* immunity, T cells are the predominant inducers of such immunity. The essential role of T cells in protection against *M. tuberculosis* infection is illustrated by the frequent occurrence of *M. tuberculosis* in AIDS patients, due to the depletion of CD4⁺ T cells associated with human immunodeficiency virus (HIV) infection. *Mycobacterium* reactive CD4⁺ T cells have been shown to be potent producers of gamma-interferon (IFN- γ), which, in turn, has been shown to trigger the antimycobacterial effects of macrophages in mice. While the role of IFN- γ in humans is

to inhibit *M. tuberculosis* infection. Furthermore, it is known that IFN- γ stimulates human macrophages to make 1,25-dihydroxy-vitamin D₃. Similarly, IL-12 has been shown to play a role in stimulating resistance to *M. tuberculosis* infection. For a review of the immunology of *M. tuberculosis* infection see Chan and Kaufmann in
 5 *Tuberculosis: Pathogenesis, Protection and Control*, Bloom (ed.), ASM Press, Washington, DC, 1994.

Accordingly, there is a need in the art for improved vaccines and methods for preventing, treating and detecting tuberculosis. The present invention fulfills these needs and further provides other related advantages.

SUMMARY OF THE INVENTION

Briefly stated, this invention provides compounds and methods for preventing and diagnosing tuberculosis. In one aspect, polypeptides are provided comprising an immunogenic portion of a soluble *M. tuberculosis* antigen, or a variant of
 15 such an antigen that differs only in conservative substitutions and/or modifications. In one embodiment of this aspect, the soluble antigen has one of the following N-terminal sequences:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-
 Gln-Val-Val-Ala-Ala-Leu (SEQ ID No. 120)
- (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-
 Ser (SEQ ID No. 121)
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-
 Ala-Lys-Glu-Gly-Arg (SEQ ID No. 122)
- (d) Tyr-Lys-Thr-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Ile-Gly-
 Pro (SEQ ID No. 123)
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val
 (SEQ ID No. 124)
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro (SEQ ID

- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly; (SEQ ID No. 127)
- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn; (SEQ ID No. 128)
- (j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID No. 134)
- (k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID No. 135) or
- (l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 136)

wherein Xaa may be any amino acid.

In a related aspect, polypeptides are provided comprising an immunogenic portion of an *M. tuberculosis* antigen, or a variant of such an antigen that differs only in conservative substitutions and/or modifications, the antigen having one of the following N-terminal sequences:

- (m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID No. 137) or
- (n) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID No. 129)

wherein Xaa may be any amino acid.

In another embodiment, the soluble *M. tuberculosis* antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID Nos. 1, 2, 4-10, 13-25, 52, 99 and 101, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos. 1, 2, 4-10, 13-25, 52, 99 and 101 or a complement thereof under moderately stringent conditions.

In a related aspect, the polypeptides comprise an immunogenic portion of

amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID Nos.: 26-51, 138, 139, 163-183, 201, 240, 242-247, 253-256, 295-298, 309, 316, 318-320, 322, 324, 328, 339, 333, 335, 337, 339 and 341, the complements of said sequences, and DNA sequences that hybridize to a sequence
5 recited in SEQ ID Nos.: 26-51, 138, 139, 163-183, 201, 240, 242-247, 253-256, 295-298, 309, 316, 318-320, 322, 324, 328, 329, 333, 335, 337, 339 and 341 or a complement thereof under moderately stringent conditions.

In related aspects, DNA sequences encoding the above polypeptides, expression vectors comprising these DNA sequences and host cells transformed or
10 transected with such expression vectors are also provided.

In another aspect, the present invention provides fusion proteins comprising a first and a second inventive polypeptide or, alternatively, an inventive polypeptide and a known *M. tuberculosis* antigen.

Within other aspects, the present invention provides pharmaceutical
15 compositions that comprise one or more of the above polypeptides, or a DNA molecule encoding such polypeptides, and a physiologically acceptable carrier. The invention also provides vaccines comprising one or more of the polypeptides as described above and a non-specific immune response enhancer, together with vaccines comprising one or more DNA sequences encoding such polypeptides and a non-specific immune
20 response enhancer.

In yet another aspect, methods are provided for inducing protective immunity in a patient, comprising administering to a patient an effective amount of one or more of the above polypeptides.

In further aspects of this invention, methods and diagnostic kits are
25 provided for detecting tuberculosis in a patient. The methods comprise contacting dermal cells of a patient with one or more of the above polypeptides and detecting an immune response on the patient's skin. The diagnostic kits comprise one or more of the above polypeptides in combination with an apparatus sufficient to contact the

In yet other aspects, methods are provided for detecting tuberculosis in a patient, such methods comprising contacting dermal cells of a patient with one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347; and detecting an immune response on the patient's skin. Diagnostic kits for use in such methods are also provided.

These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE IDENTIFIERS

Figure 1A and B illustrate the stimulation of proliferation and interferon- γ production in T cells derived from a first and a second *M. tuberculosis*-immune donor, respectively, by the 14 Kd, 20 Kd and 26 Kd antigens described in Example 1.

Figure 2 illustrates the stimulation of proliferation and interferon- γ production in T cells derived from an *M. tuberculosis*-immune individual by the two representative polypeptides TbRa3 and TbRa9.

Figures 3A-D illustrate the reactivity of antisera raised against secretory *M. tuberculosis* proteins, the known *M. tuberculosis* antigen 85b and the inventive antigens Tb38-1 and TbH-9, respectively, with *M. tuberculosis* lysate (lane 2), *M. tuberculosis* secretory proteins (lane 3), recombinant Tb38-1 (lane 4), recombinant TbH-9 (lane 5) and recombinant 85b (lane 6).

Figure 4B illustrates the stimulation of interferon- γ production in a TbH-9-specific T cell clone by secretory *M. tuberculosis* proteins, PPD and recombinant TbH-9.

Figures 5A and B illustrate the stimulation of proliferation and interferon- γ production in TbH9-specific T cells by the fusion protein TbH9-Tb38-1.

Figures 6A and B illustrate the stimulation of proliferation and interferon- γ production in Tb38-1-specific T cells by the fusion protein TbH9-Tb38-1.

Figures 7A and B illustrate the stimulation of proliferation and interferon- γ production in T cells previously shown to respond to both TbH-9 and Tb38-1 by the fusion protein TbH9-Tb38-1.

Figures 8A and B illustrate the stimulation of proliferation and interferon- γ production in T cells derived from a first *M. tuberculosis*-immune individual by the representative polypeptides XP-1, RDIF6, RDIF8, RDIF10 and RDIF11.

Figures 9A and B illustrate the stimulation of proliferation and interferon- γ production in T cells derived from a second *M. tuberculosis*-immune individual by the representative polypeptides XP-1, RDIF6, RDIF8, RDIF10 and RDIF11.

SEQ. ID NO. 1 is the DNA sequence of TbRa1.

SEQ. ID NO. 2 is the DNA sequence of TbRa10.

SEQ. ID NO. 3 is the DNA sequence of TbRa11.

SEQ. ID NO. 4 is the DNA sequence of TbRa12.

SEQ. ID NO. 5 is the DNA sequence of TbRa13.

SEQ. ID NO. 6 is the DNA sequence of TbRa16.

SEQ. ID NO. 7 is the DNA sequence of TbRa17.

SEQ. ID NO. 8 is the DNA sequence of TbRa18.

SEQ. ID NO. 9 is the DNA sequence of TbRa19.

SEQ. ID NO. 10 is the DNA sequence of TbRa24.

SEQ. ID NO. 13 is the DNA sequence of TbRa29.
SEQ. ID NO. 14 is the DNA sequence of TbRa2A.
SEQ. ID NO. 15 is the DNA sequence of TbRa3.
SEQ. ID NO. 16 is the DNA sequence of TbRa32.
5 SEQ. ID NO. 17 is the DNA sequence of TbRa35.
SEQ. ID NO. 18 is the DNA sequence of TbRa36.
SEQ. ID NO. 19 is the DNA sequence of TbRa4.
SEQ. ID NO. 20 is the DNA sequence of TbRa9.
SEQ. ID NO. 21 is the DNA sequence of TbRaB.
10 SEQ. ID NO. 22 is the DNA sequence of TbRaC.
SEQ. ID NO. 23 is the DNA sequence of TbRaD.
SEQ. ID NO. 24 is the DNA sequence of YWCPG.
SEQ. ID NO. 25 is the DNA sequence of AAMK.
SEQ. ID NO. 26 is the DNA sequence of TbL-23.
15 SEQ. ID NO. 27 is the DNA sequence of TbL-24.
SEQ. ID NO. 28 is the DNA sequence of TbL-25.
SEQ. ID NO. 29 is the DNA sequence of TbL-28.
SEQ. ID NO. 30 is the DNA sequence of TbL-29.
SEQ. ID NO. 31 is the DNA sequence of TbH-5.
20 SEQ. ID NO. 32 is the DNA sequence of TbH-8.
SEQ. ID NO. 33 is the DNA sequence of TbH-9.
SEQ. ID NO. 34 is the DNA sequence of TbM-1.
SEQ. ID NO. 35 is the DNA sequence of TbM-3.
SEQ. ID NO. 36 is the DNA sequence of TbM-6.
25 SEQ. ID NO. 37 is the DNA sequence of TbM-7.
SEQ. ID NO. 38 is the DNA sequence of TbM-9.
SEQ. ID NO. 39 is the DNA sequence of TbM-12.
SEQ. ID NO. 40 is the DNA sequence of TbM-17.

- SEQ. ID NO. 43 is the DNA sequence of TbH-4.
SEQ. ID NO. 44 is the DNA sequence of TbH-4-FWD.
SEQ. ID NO. 45 is the DNA sequence of TbH-12.
SEQ. ID NO. 46 is the DNA sequence of Tb38-1.
5 SEQ. ID NO. 47 is the DNA sequence of Tb38-4.
SEQ. ID NO. 48 is the DNA sequence of TbL-17.
SEQ. ID NO. 49 is the DNA sequence of TbL-20.
SEQ. ID NO. 50 is the DNA sequence of TbL-21.
SEQ. ID NO. 51 is the DNA sequence of TbH-16.
10 SEQ. ID NO. 52 is the DNA sequence of DPEP.
SEQ. ID NO. 53 is the deduced amino acid sequence of DPEP.
SEQ. ID NO. 54 is the protein sequence of DPV N-terminal Antigen.
SEQ. ID NO. 55 is the protein sequence of AVGS N-terminal Antigen.
SEQ. ID NO. 56 is the protein sequence of AAMK N-terminal Antigen.
15 SEQ. ID NO. 57 is the protein sequence of YYWC N-terminal Antigen.
SEQ. ID NO. 58 is the protein sequence of DIGS N-terminal Antigen.
SEQ. ID NO. 59 is the protein sequence of AEES N-terminal Antigen.
SEQ. ID NO. 60 is the protein sequence of DPEP N-terminal Antigen.
SEQ. ID NO. 61 is the protein sequence of APK7 N-terminal Antigen.
20 SEQ. ID NO. 62 is the protein sequence of DPAS N-terminal Antigen.
SEQ. ID NO. 63 is the deduced amino acid sequence of TbRa1.
SEQ. ID NO. 64 is the deduced amino acid sequence of TbRa10.
SEQ. ID NO. 65 is the deduced amino acid sequence of TbRa11.
SEQ. ID NO. 66 is the deduced amino acid sequence of TbRa12.
25 SEQ. ID NO. 67 is the deduced amino acid sequence of TbRa13.
SEQ. ID NO. 68 is the deduced amino acid sequence of TbRa16.
SEQ. ID NO. 69 is the deduced amino acid sequence of TbRa17.
SEQ. ID NO. 70 is the deduced amino acid sequence of TbRa18.

SEQ. ID NO. 73 is the deduced amino acid sequence of TbRa26.
SEQ. ID NO. 74 is the deduced amino acid sequence of TbRa28.
SEQ. ID NO. 75 is the deduced amino acid sequence of TbRa29.
SEQ. ID NO. 76 is the deduced amino acid sequence of TbRa2A.
5 SEQ. ID NO. 77 is the deduced amino acid sequence of TbRa3.
SEQ. ID NO. 78 is the deduced amino acid sequence of TbRa32.
SEQ. ID NO. 79 is the deduced amino acid sequence of TbRa35.
SEQ. ID NO. 80 is the deduced amino acid sequence of TbRa36.
SEQ. ID NO. 81 is the deduced amino acid sequence of TbRa4.
10 SEQ. ID NO. 82 is the deduced amino acid sequence of TbRa9.
SEQ. ID NO. 83 is the deduced amino acid sequence of TbRaB.
SEQ. ID NO. 84 is the deduced amino acid sequence of TbRaC.
SEQ. ID NO. 85 is the deduced amino acid sequence of TbRaD.
SEQ. ID NO. 86 is the deduced amino acid sequence of YYWCPG.
15 SEQ. ID NO. 87 is the deduced amino acid sequence of TbAAMK.
SEQ. ID NO. 88 is the deduced amino acid sequence of Tb38-1.
SEQ. ID NO. 89 is the deduced amino acid sequence of TbH-4.
SEQ. ID NO. 90 is the deduced amino acid sequence of TbH-8.
SEQ. ID NO. 91 is the deduced amino acid sequence of TbH-9.
20 SEQ. ID NO. 92 is the deduced amino acid sequence of TbH-12.
SEQ. ID NO. 93 is the amino acid sequence of Tb38-1 Peptide 1.
SEQ. ID NO. 94 is the amino acid sequence of Tb38-1 Peptide 2.
SEQ. ID NO. 95 is the amino acid sequence of Tb38-1 Peptide 3.
SEQ. ID NO. 96 is the amino acid sequence of Tb38-1 Peptide 4.
25 SEQ. ID NO. 97 is the amino acid sequence of Tb38-1 Peptide 5.
SEQ. ID NO. 98 is the amino acid sequence of Tb38-1 Peptide 6.
SEQ. ID NO. 99 is the DNA sequence of DPAS.
SEQ. ID NO. 100 is the deduced amino acid sequence of DPAS.

SEQ. ID NO. 103 is the DNA sequence of ESAT-6.

SEQ. ID NO. 104 is the deduced amino acid sequence of ESAT-6.

SEQ. ID NO. 105 is the DNA sequence of TbH-8-2.

SEQ. ID NO. 106 is the DNA sequence of TbH-9FL.

5 SEQ. ID NO. 107 is the deduced amino acid sequence of TbH-9FL.

SEQ. ID NO. 108 is the DNA sequence of TbH-9-1.

SEQ. ID NO. 109 is the deduced amino acid sequence of TbH-9-1.

SEQ. ID NO. 110 is the DNA sequence of TbH-9-4.

SEQ. ID NO. 111 is the deduced amino acid sequence of TbH-9-4.

10 SEQ. ID NO. 112 is the DNA sequence of Tb38-1F2 IN.

SEQ. ID NO. 113 is the DNA sequence of Tb38-2F2 RP.

SEQ. ID NO. 114 is the deduced amino acid sequence of Tb37-FL.

SEQ. ID NO. 115 is the deduced amino acid sequence of Tb38-IN.

SEQ. ID NO. 116 is the DNA sequence of Tb38-1F3.

15 SEQ. ID NO. 117 is the deduced amino acid sequence of Tb38-1F3.

SEQ. ID NO. 118 is the DNA sequence of Tb38-1F5.

SEQ. ID NO. 119 is the DNA sequence of Tb38-1F6.

SEQ. ID NO. 120 is the deduced N-terminal amino acid sequence of DPV.

SEQ. ID NO. 121 is the deduced N-terminal amino acid sequence of AVGS.

20 SEQ. ID NO. 122 is the deduced N-terminal amino acid sequence of AAMK.

SEQ. ID NO. 123 is the deduced N-terminal amino acid sequence of YYWC.

SEQ. ID NO. 124 is the deduced N-terminal amino acid sequence of DIGS.

SEQ. ID NO. 125 is the deduced N-terminal amino acid sequence of AEES.

SEQ. ID NO. 126 is the deduced N-terminal amino acid sequence of DPEP.

25 SEQ. ID NO. 127 is the deduced N-terminal amino acid sequence of APKT.

SEQ. ID NO. 128 is the deduced amino acid sequence of DPAS.

SEQ. ID NO. 129 is the protein sequence of DPPD N-terminal Antigen.

SEQ. ID NO. 130-133 are the protein sequences of four DPPD cyanogen

SEQ ID NO. 135 is the N-terminal protein sequence of AGD antigen.

SEQ ID NO. 136 is the N-terminal protein sequence of APE antigen.

SEQ ID NO. 137 is the N-terminal protein sequence of XYI antigen.

SEQ ID NO. 138 is the DNA sequence of TbH-29.

5 SEQ ID NO. 139 is the DNA sequence of TbH-30.

SEQ ID NO. 140 is the DNA sequence of TbH-32.

SEQ ID NO. 141 is the DNA sequence of TbH-33.

SEQ ID NO. 142 is the predicted amino acid sequence of TbH-29.

SEQ ID NO. 143 is the predicted amino acid sequence of TbH-30.

10 SEQ ID NO. 144 is the predicted amino acid sequence of TbH-32.

SEQ ID NO. 145 is the predicted amino acid sequence of TbH-33.

SEQ ID NO: 146-151 are PCR primers used in the preparation of a fusion protein containing TbRa3, 38 kD and Tb38-1.

15 SEQ ID NO: 152 is the DNA sequence of the fusion protein containing TbRa3, 38 kD and Tb38-1.

SEQ ID NO: 153 is the amino acid sequence of the fusion protein containing TbRa3, 38 kD and Tb38-1.

SEQ ID NO: 154 is the DNA sequence of the *M. tuberculosis* antigen 38 kD.

20 SEQ ID NO: 155 is the amino acid sequence of the *M. tuberculosis* antigen 38 kD.

SEQ ID NO: 156 is the DNA sequence of XP14.

SEQ ID NO: 157 is the DNA sequence of XP24.

SEQ ID NO: 158 is the DNA sequence of XP31.

SEQ ID NO: 159 is the 5' DNA sequence of XP32.

25 SEQ ID NO: 160 is the 3' DNA sequence of XP32.

SEQ ID NO: 161 is the predicted amino acid sequence of XP14.

SEQ ID NO: 162 is the predicted amino acid sequence encoded by the reverse complement of XP14.

SEQ ID NO: 163 is the

- SEQ ID NO: 165 is the 5' DNA sequence of XP4.
SEQ ID NO: 166 is the 5' DNA sequence of XP5.
SEQ ID NO: 167 is the 5' DNA sequence of XP17.
SEQ ID NO: 168 is the 5' DNA sequence of XP30.
5 SEQ ID NO: 169 is the 5' DNA sequence of XP2.
SEQ ID NO: 170 is the 3' DNA sequence of XP2.
SEQ ID NO: 171 is the 5' DNA sequence of XP3.
SEQ ID NO: 172 is the 3' DNA sequence of XP3.
SEQ ID NO: 173 is the 5' DNA sequence of XP6.
10 SEQ ID NO: 174 is the 3' DNA sequence of XP6.
SEQ ID NO: 175 is the 5' DNA sequence of XP18.
SEQ ID NO: 176 is the 3' DNA sequence of XP18.
SEQ ID NO: 177 is the 5' DNA sequence of XP19.
SEQ ID NO: 178 is the 3' DNA sequence of XP19.
15 SEQ ID NO: 179 is the 5' DNA sequence of XP22.
SEQ ID NO: 180 is the 3' DNA sequence of XP22.
SEQ ID NO: 181 is the 5' DNA sequence of XP25.
SEQ ID NO: 182 is the 3' DNA sequence of XP25.
SEQ ID NO: 183 is the full-length DNA sequence of TbH4-XP1.
20 SEQ ID NO: 184 is the predicted amino acid sequence of TbH4-XP1.
SEQ ID NO: 185 is the predicted amino acid sequence encoded by the reverse
complement of TbH4-XP1.
SEQ ID NO: 186 is a first predicted amino acid sequence encoded by XP36.
SEQ ID NO: 187 is a second predicted amino acid sequence encoded by XP36.
25 SEQ ID NO: 188 is the predicted amino acid sequence encoded by the reverse
complement of XP36.
SEQ ID NO: 189 is the DNA sequence of RDIF2.
SEQ ID NO: 190 is the DNA sequence of RDIF5.

SEQ ID NO: 193 is the DNA sequence of RDIF11.
SEQ ID NO: 194 is the predicted amino acid sequence of RDIF2.
SEQ ID NO: 195 is the predicted amino acid sequence of RDIF5.
SEQ ID NO: 196 is the predicted amino acid sequence of RDIF8.
5 SEQ ID NO: 197 is the predicted amino acid sequence of RDIF10.
SEQ ID NO: 198 is the predicted amino acid sequence of RDIF11.
SEQ ID NO: 199 is the 5' DNA sequence of RDIF12.
SEQ ID NO: 200 is the 3' DNA sequence of RDIF12.
SEQ ID NO: 201 is the DNA sequence of RDIF7.
10 SEQ ID NO: 202 is the predicted amino acid sequence of RDIF7.
SEQ ID NO: 203 is the DNA sequence of DIF2-1.
SEQ ID NO: 204 is the predicted amino acid sequence of DIF2-1.
SEQ ID NO: 205-212 are PCR primers used in the preparation of a fusion
protein containing TbRa3, 38 kD, Tb38-1 and DPEP (hereinafter referred to as
15 TbF-2).
SEQ ID NO: 213 is the DNA sequence of the fusion protein TbF-2.
SEQ ID NO: 214 is the amino acid sequence of the fusion protein TbF-2.
SEQ ID NO: 215 is the 5' DNA sequence of MO-1.
SEQ ID NO: 216 is the 5' DNA sequence for MO-2.
20 SEQ ID NO: 217 is the 5' DNA sequence for MO-4.
SEQ ID NO: 218 is the 5' DNA sequence for MO-8.
SEQ ID NO: 219 is the 5' DNA sequence for MO-9.
SEQ ID NO: 220 is the 5' DNA sequence for MO-26.
SEQ ID NO: 221 is the 5' DNA sequence for MO-28.
25 SEQ ID NO: 222 is the 5' DNA sequence for MO-29.
SEQ ID NO: 223 is the 5' DNA sequence for MO-30.
SEQ ID NO: 224 is the 5' DNA sequence for MO-34.
SEQ ID NO: 225 is the 5' DNA sequence for MO-35.

- SEQ ID NO: 228 is the predicted amino acid sequence for MO-4.
SEQ ID NO: 229 is the predicted amino acid sequence for MO-8.
SEQ ID NO: 230 is the predicted amino acid sequence for MO-9.
SEQ ID NO: 231 is the predicted amino acid sequence for MO-26.
5 SEQ ID NO: 232 is the predicted amino acid sequence for MO-28.
SEQ ID NO: 233 is the predicted amino acid sequence for MO-29.
SEQ ID NO: 234 is the predicted amino acid sequence for MO-30.
SEQ ID NO: 235 is the predicted amino acid sequence for MO-34.
SEQ ID NO: 236 is the predicted amino acid sequence for MO-35.
10 SEQ ID NO: 237 is the determined DNA sequence for MO-10.
SEQ ID NO: 238 is the predicted amino acid sequence for MO-10.
SEQ ID NO: 239 is the 3' DNA sequence for MO-27.
SEQ ID NO: 240 is the full-length DNA sequence for DPPD.
SEQ ID NO: 241 is the predicted full-length amino acid sequence for DPPD.
15 SEQ ID NO: 242 is the determined 5' cDNA sequence for LSER-10.
SEQ ID NO: 243 is the determined 5' cDNA sequence for LSER-11.
SEQ ID NO: 244 is the determined 5' cDNA sequence for LSER-12.
SEQ ID NO: 245 is the determined 5' cDNA sequence for LSER-13.
SEQ ID NO: 246 is the determined 5' cDNA sequence for LSER-16.
20 SEQ ID NO: 247 is the determined 5' cDNA sequence for LSER-25.
SEQ ID NO: 248 is the predicted amino acid sequence for LSER-10.
SEQ ID NO: 249 is the predicted amino acid sequence for LSER-12.
SEQ ID NO: 250 is the predicted amino acid sequence for LSER-13.
SEQ ID NO: 251 is the predicted amino acid sequence for LSER-16.
25 SEQ ID NO: 252 is the predicted amino acid sequence for LSER-25.
SEQ ID NO: 253 is the determined cDNA sequence for LSER-18.
SEQ ID NO: 254 is the determined cDNA sequence for LSER-23.
SEQ ID NO: 255 is the determined cDNA sequence for LSER-24.

SEQ ID NO: 258 is the predicted amino acid sequence for LSER-23
SEQ ID NO: 259 is the predicted amino acid sequence for LSER-24
SEQ ID NO: 260 is the predicted amino acid sequence for LSER-27
SEQ ID NO: 261 is the determined 5' cDNA sequence for LSER-1
5 SEQ ID NO: 262 is the determined 5' cDNA sequence for LSER-3
SEQ ID NO: 263 is the determined 5' cDNA sequence for LSER-4
SEQ ID NO: 264 is the determined 5' cDNA sequence for LSER-5
SEQ ID NO: 265 is the determined 5' cDNA sequence for LSER-6
SEQ ID NO: 266 is the determined 5' cDNA sequence for LSER-8
10 SEQ ID NO: 267 is the determined 5' cDNA sequence for LSER-14
SEQ ID NO: 268 is the determined 5' cDNA sequence for LSER-15
SEQ ID NO: 269 is the determined 5' cDNA sequence for LSER-17
SEQ ID NO: 270 is the determined 5' cDNA sequence for LSER-19
SEQ ID NO: 271 is the determined 5' cDNA sequence for LSER-20
15 SEQ ID NO: 272 is the determined 5' cDNA sequence for LSER-22
SEQ ID NO: 273 is the determined 5' cDNA sequence for LSER-26
SEQ ID NO: 274 is the determined 5' cDNA sequence for LSER-28
SEQ ID NO: 275 is the determined 5' cDNA sequence for LSER-29
SEQ ID NO: 276 is the determined 5' cDNA sequence for LSER-30
20 SEQ ID NO: 277 is the predicted amino acid sequence for LSER-1
SEQ ID NO: 278 is the predicted amino acid sequence for LSER-3
SEQ ID NO: 279 is the predicted amino acid sequence for LSER-5
SEQ ID NO: 280 is the predicted amino acid sequence for LSER-6
SEQ ID NO: 281 is the predicted amino acid sequence for LSER-8
25 SEQ ID NO: 282 is the predicted amino acid sequence for LSER-14
SEQ ID NO: 283 is the predicted amino acid sequence for LSER-15
SEQ ID NO: 284 is the predicted amino acid sequence for LSER-17
SEQ ID NO: 285 is the predicted amino acid sequence for LSER-19

SEQ ID NO: 288 is the predicted amino acid sequence for LSER-26

SEQ ID NO: 289 is the predicted amino acid sequence for LSER-28

SEQ ID NO: 290 is the predicted amino acid sequence for LSER-29

SEQ ID NO: 291 is the predicted amino acid sequence for LSER-30

5 SEQ ID NO: 292 is the determined cDNA sequence for LSER-9

SEQ ID NO: 293 is the determined cDNA sequence for the reverse complement of LSER-6

SEQ ID NO: 294 is the predicted amino acid sequence for the reverse complement of LSER-6

10 SEQ ID NO: 295 is the determined 5' cDNA sequence for MO-12

SEQ ID NO: 296 is the determined 5' cDNA sequence for MO-13

SEQ ID NO: 297 is the determined 5' cDNA sequence for MO-19

SEQ ID NO: 298 is the determined 5' cDNA sequence for MO-39

SEQ ID NO: 299 is the predicted amino acid sequence for MO-12

15 SEQ ID NO: 300 is the predicted amino acid sequence for MO-13

SEQ ID NO: 301 is the predicted amino acid sequence for MO-19

SEQ ID NO: 302 is the predicted amino acid sequence for MO-39

SEQ ID NO: 303 is the determined 5' cDNA sequence for Erdsn-1

SEQ ID NO: 304 is the determined 5' cDNA sequence for Erdsn-2

20 SEQ ID NO: 305 is the determined 5' cDNA sequence for Erdsn-4

SEQ ID NO: 306 is the determined 5' cDNA sequence for Erdsn-5

SEQ ID NO: 307 is the determined 5' cDNA sequence for Erdsn-6

SEQ ID NO: 308 is the determined 5' cDNA sequence for Erdsn-7

SEQ ID NO: 309 is the determined 5' cDNA sequence for Erdsn-8

25 SEQ ID NO: 310 is the determined 5' cDNA sequence for Erdsn-9

SEQ ID NO: 311 is the determined 5' cDNA sequence for Erdsn-10

SEQ ID NO: 312 is the determined 5' cDNA sequence for Erdsn-12

SEQ ID NO: 313 is the determined 5' cDNA sequence for Erdsn-13

SEQ ID NO: 314 is the determined 5' cDNA sequence for Erdsn-14

SEQ ID NO: 316 is the determined 5' cDNA sequence for Erdsn-16

SEQ ID NO: 317 is the determined 5' cDNA sequence for Erdsn-17

SEQ ID NO: 318 is the determined 5' cDNA sequence for Erdsn-18

SEQ ID NO: 319 is the determined 5' cDNA sequence for Erdsn-21

5 SEQ ID NO: 320 is the determined 5' cDNA sequence for Erdsn-22

SEQ ID NO: 321 is the determined 5' cDNA sequence for Erdsn-23

SEQ ID NO: 322 is the determined 5' cDNA sequence for Erdsn-25

SEQ ID NO: 323 is the determined 3' cDNA sequence for Erdsn-1

SEQ ID NO: 324 is the determined 3' cDNA sequence for Erdsn-2

10 SEQ ID NO: 325 is the determined 3' cDNA sequence for Erdsn-4

SEQ ID NO: 326 is the determined 3' cDNA sequence for Erdsn-5

SEQ ID NO: 327 is the determined 3' cDNA sequence for Erdsn-7

SEQ ID NO: 328 is the determined 3' cDNA sequence for Erdsn-8

SEQ ID NO: 329 is the determined 3' cDNA sequence for Erdsn-9

15 SEQ ID NO: 330 is the determined 3' cDNA sequence for Erdsn-10

SEQ ID NO: 331 is the determined 3' cDNA sequence for Erdsn-12

SEQ ID NO: 332 is the determined 3' cDNA sequence for Erdsn-13

SEQ ID NO: 333 is the determined 3' cDNA sequence for Erdsn-14

SEQ ID NO: 334 is the determined 3' cDNA sequence for Erdsn-15

20 SEQ ID NO: 335 is the determined 3' cDNA sequence for Erdsn-16

SEQ ID NO: 336 is the determined 3' cDNA sequence for Erdsn-17

SEQ ID NO: 337 is the determined 3' cDNA sequence for Erdsn-18

SEQ ID NO: 338 is the determined 3' cDNA sequence for Erdsn-21

SEQ ID NO: 339 is the determined 3' cDNA sequence for Erdsn-22

25 SEQ ID NO: 340 is the determined 3' cDNA sequence for Erdsn-23

SEQ ID NO: 341 is the determined 3' cDNA sequence for Erdsn-25

SEQ ID NO: 342 is the determined cDNA sequence for Erdsn-24

SEQ ID NO: 343 is the determined amino acid sequence for a *M. tuberculosis*

85b premature hemolysin

- SEQ ID NO: 345 is a determined amino acid sequence for spot 2
SEQ ID NO: 346 is a determined amino acid sequence for spot 2
SEQ ID NO: 347 is the determined amino acid seq for spot 4
SEQ ID NO: 348 is the sequence of primer PDM-157
5 SEQ ID NO: 349 is the sequence of primer PDM-160
SEQ ID NO: 350 is the DNA sequence of the fusion protein TbF-6
SEQ ID NO: 351 is the amino acid sequence of fusion protein TbF-6
SEQ ID NO: 352 is the sequence of primer PDM-176
SEQ ID NO: 353 is the sequence of primer PDM-175
10 SEQ ID NO: 354 is the DNA sequence of the fusion protein TbF-8
SEQ ID NO: 355 is the amino acid sequence of the fusion protein TbF-8

DETAILED DESCRIPTION OF THE INVENTION

As noted above, the present invention is generally directed to
15 compositions and methods for preventing, treating and diagnosing tuberculosis. The
compositions of the subject invention include polypeptides that comprise at least one
immunogenic portion of a *M. tuberculosis* antigen, or a variant of such an antigen that
differs only in conservative substitutions and/or modifications. Polypeptides within the
scope of the present invention include, but are not limited to, immunogenic soluble
20 *M. tuberculosis* antigens. A "soluble *M. tuberculosis* antigen" is a protein of
M. tuberculosis origin that is present in *M. tuberculosis* culture filtrate. As used herein,
the term "polypeptide" encompasses amino acid chains of any length, including full
length proteins (i.e., antigens), wherein the amino acid residues are linked by covalent
peptide bonds. Thus, a polypeptide comprising an immunogenic portion of one of the
25 above antigens may consist entirely of the immunogenic portion, or may contain
additional sequences. The additional sequences may be derived from the native
M. tuberculosis antigen or may be heterologous, and such sequences may (but need not)
be immunogenic.

of such antigens) are capable of stimulating cell proliferation, interleukin-12 production and/or interferon- γ production in biological samples comprising one or more cells selected from the group of T cells, NK cells, B cells and macrophages, where the cells are derived from an *M. tuberculosis*-immune individual. Polypeptides comprising at least an immunogenic portion of one or more *M. tuberculosis* antigens may generally be used to detect tuberculosis or to induce protective immunity against tuberculosis in a patient.

The compositions and methods of the present invention also encompass variants of the above polypeptides and DNA molecules. A polypeptide "variant," as used herein, is a polypeptide that differs from the recited polypeptide only in conservative substitutions and/or modifications, such that the therapeutic, antigenic and/or immunogenic properties of the polypeptide are retained. Polypeptide variants preferably exhibit at least about 70%, more preferably at least about 90% and most preferably at least about 95% identity to the identified polypeptides. For polypeptides with immunoreactive properties, variants may, alternatively, be identified by modifying the amino acid sequence of one of the above polypeptides, and evaluating the immunoreactivity of the modified polypeptide. For polypeptides useful for the generation of diagnostic binding agents, a variant may be identified by evaluating a modified polypeptide for the ability to generate antibodies that detect the presence or absence of tuberculosis. Such modified sequences may be prepared and tested using, for example, the representative procedures described herein.

As used herein, a "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydropathic nature of the polypeptide to be substantially unchanged. In general, the following groups of amino acids represent conservative changes: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his.

antigenic properties, secondary structure and hydrophobic nature of the polypeptide. For example, a polypeptide may be conjugated to a signal (or leader) sequence at the N-terminal end of the protein which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (e.g., poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

A nucleotide "variant" is a sequence that differs from the recited nucleotide sequence in having one or more nucleotide deletions, substitutions or additions. Such modifications may be readily introduced using standard mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis as taught, for example, by Adelman et al. (*DNA*, 2:183, 1983). Nucleotide variants may be naturally occurring allelic variants, or non-naturally occurring variants. Variant nucleotide sequences preferably exhibit at least about 70%, more preferably at least about 80% and most preferably at least about 90% identity to the recited sequence. Such variant nucleotide sequences will generally hybridize to the recited nucleotide sequence under stringent conditions. As used herein, "stringent conditions" refers to prewashing in a solution of 6X SSC, 0.2% SDS; hybridizing at 65 °C, 6X SSC, 0.2% SDS overnight; followed by two washes of 30 minutes each in 1X SSC, 0.1% SDS at 65 °C and two washes of 30 minutes each in 0.2X SSC, 0.1% SDS at 65 °C.

In a related aspect, combination polypeptides are disclosed. A "combination polypeptide" is a polypeptide comprising at least one of the above immunogenic portions and one or more additional immunogenic *M. tuberculosis* sequences, which are joined via a peptide linkage into a single amino acid chain. The sequences may be joined directly (*i.e.*, with no intervening amino acids) or may be joined by way of a linker sequence (e.g., Gly-Cys-Gly) that does not significantly diminish the immunogenic properties of the component polypeptides.

In general, *M. tuberculosis* antigens, and DNA sequences, are disclosed.

those of ordinary skill in the art, including anion-exchange and reverse phase chromatography. Purified antigens are then evaluated for their ability to elicit an appropriate immune response (e.g., cellular) using, for example, the representative methods described herein. Immunogenic antigens may then be partially sequenced
5 using techniques such as traditional Edman chemistry. See Edman and Berg, *Eur. J. Biochem.* 80:116-132, 1967.

Immunogenic antigens may also be produced recombinantly using a DNA sequence that encodes the antigen, which has been inserted into an expression vector and expressed in an appropriate host. DNA molecules encoding soluble antigens
10 may be isolated by screening an appropriate *M. tuberculosis* expression library with anti-sera (e.g., rabbit) raised specifically against soluble *M. tuberculosis* antigens. DNA sequences encoding antigens that may or may not be soluble may be identified by screening an appropriate *M. tuberculosis* genomic or cDNA expression library with sera obtained from patients infected with *M. tuberculosis*. Such screens may generally be
15 performed using techniques well known to those of ordinary skill in the art, such as those described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989.

DNA sequences encoding soluble antigens may also be obtained by screening an appropriate *M. tuberculosis* cDNA or genomic DNA library for DNA
20 sequences that hybridize to degenerate oligonucleotides derived from partial amino acid sequences of isolated soluble antigens. Degenerate oligonucleotide sequences for use in such a screen may be designed and synthesized, and the screen may be performed, as described (for example) in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989 (and references cited therein). Polymerase chain reaction (PCR) may also be employed, using the above
25 oligonucleotides in methods well known in the art, to isolate a nucleic acid probe from a cDNA or genomic library. The library screen may then be performed using the isolated probe.

Alternatively, genomic or cDNA libraries may be screened using

lines or clones derived from one or more *M. tuberculosis*-immune individuals. In general, PBMCs and/or T cells for use in such screens may be prepared as described below. Direct library screens may generally be performed by assaying pools of expressed recombinant proteins for the ability to induce proliferation and/or interferon- γ production in T cells derived from an *M. tuberculosis*-immune individual. Alternatively, potential T cell antigens may be first selected based on antibody reactivity, as described above.

Regardless of the method of preparation, the antigens (and immunogenic portions thereof) described herein (which may or may not be soluble) have the ability to induce an immunogenic response. More specifically, the antigens have the ability to induce proliferation and/or cytokine production (i.e., interferon- γ and/or interleukin-12 production) in T cells, NK cells, B cells and/or macrophages derived from an *M. tuberculosis*-immune individual. The selection of cell type for use in evaluating an immunogenic response to a antigen will, of course, depend on the desired response. For example, interleukin-12 production is most readily evaluated using preparations containing B cells and/or macrophages. An *M. tuberculosis*-immune individual is one who is considered to be resistant to the development of tuberculosis by virtue of having mounted an effective T cell response to *M. tuberculosis* (i.e., substantially free of disease symptoms). Such individuals may be identified based on a strongly positive (i.e., greater than about 10 mm diameter induration) intradermal skin test response to tuberculosis proteins (PPD) and an absence of any signs or symptoms of tuberculosis disease. T cells, NK cells, B cells and macrophages derived from *M. tuberculosis*-immune individuals may be prepared using methods known to those of ordinary skill in the art. For example, a preparation of PBMCs (i.e., peripheral blood mononuclear cells) may be employed without further separation of component cells. PBMCs may generally be prepared, for example, using density centrifugation through FicollTM (Winthrop Laboratories, NY). T cells for use in the assays described herein may also be purified directly from PBMCs. Alternatively, an enriched T cell line reactive against mycobacterial proteins is available.

PBMCs from *M. tuberculosis*-immune individuals with mycobacterial proteins for a period of 2-4 weeks. This allows expansion of only the mycobacterial protein-specific T cells, resulting in a line composed solely of such cells. These cells may then be cloned and tested with individual proteins, using methods known to those of ordinary skill in the art, to more accurately define individual T cell specificity. In general, antigens that test positive in assays for proliferation and/or cytokine production (*i.e.*, interferon- γ and/or interleukin-12 production) performed using T cells, NK cells, B cells and/or macrophages derived from an *M. tuberculosis*-immune individual are considered immunogenic. Such assays may be performed, for example, using the representative procedures described below. Immunogenic portions of such antigens may be identified using similar assays, and may be present within the polypeptides described herein.

The ability of a polypeptide (*e.g.*, an immunogenic antigen, or a portion or other variant thereof) to induce cell proliferation is evaluated by contacting the cells (*e.g.*, T cells and/or NK cells) with the polypeptide and measuring the proliferation of the cells. In general, the amount of polypeptide that is sufficient for evaluation of about 10^5 cells ranges from about 10 ng/mL to about 100 μ g/mL and preferably is about 10 μ g/mL. The incubation of polypeptide with cells is typically performed at 37°C for about six days. Following incubation with polypeptide, the cells are assayed for a proliferative response, which may be evaluated by methods known to those of ordinary skill in the art, such as exposing cells to a pulse of radiolabeled thymidine and measuring the incorporation of label into cellular DNA. In general, a polypeptide that results in at least a three fold increase in proliferation above background (*i.e.*, the proliferation observed for cells cultured without polypeptide) is considered to be able to induce proliferation.

The ability of a polypeptide to stimulate the production of interferon- γ and/or interleukin-12 in cells may be evaluated by contacting the cells with the polypeptide and measuring the level of interferon- γ or interleukin-12 produced by the cells. In general, the amount of polypeptide that is sufficient for the evaluation of about 10^5 cells ranges from about 10 ng/mL to about 100 μ g/mL and preferably is about 10 μ g/mL.

bead or a biodegradable microsphere, such as those described in U.S. Patent Nos. 4,897,268 and 5,075,109. The incubation of polypeptide with the cells is typically performed at 37°C for about six days. Following incubation with polypeptide, the cells are assayed for interferon- γ and/or interleukin-12 (or one or more subunits thereof), which may be evaluated by methods known to those of ordinary skill in the art, such as an enzyme-linked immunosorbent assay (ELISA) or, in the case of IL-12 P70 subunit, a bioassay such as an assay measuring proliferation of T cells. In general, a polypeptide that results in the production of at least 50 pg of interferon- γ per mL of cultured supernatant (containing 10^4 - 10^5 T cells per mL) is considered able to stimulate the production of interferon- γ . A polypeptide that stimulates the production of at least 10 pg/mL of IL-12 P70 subunit, and/or at least 100 pg/mL of IL-12 P40 subunit, per 10^5 macrophages or B cells (or per 3×10^5 PBMC) is considered able to stimulate the production of IL-12.

In general, immunogenic antigens are those antigens that stimulate proliferation and/or cytokine production (i.e., interferon- γ and/or interleukin-12 production) in T cells, NK cells, B cells and/or macrophages derived from at least about 25% of *M. tuberculosis*-immune individuals. Among these immunogenic antigens, polypeptides having superior therapeutic properties may be distinguished based on the magnitude of the responses in the above assays and based on the percentage of individuals for which a response is observed. In addition, antigens having superior therapeutic properties will not stimulate proliferation and/or cytokine production *in vitro* in cells derived from more than about 25% of individuals that are not *M. tuberculosis*-immune, thereby eliminating responses that are not specifically due to *M. tuberculosis*-responsive cells. Those antigens that induce a response in a high percentage of T cell, NK cell, B cell and/or macrophage preparations from *M. tuberculosis*-immune individuals (with a low incidence of responses in cell preparations from other individuals) have superior therapeutic properties.

Antigens with superior therapeutic properties may also be identified based on their ability to stimulate a response in

for use on experimental animals are described in detail below. Efficacy may be determined based on the ability of the antigen to provide at least about a 50% reduction in bacterial numbers and/or at least about a 40% decrease in mortality following experimental infection. Suitable experimental animals include mice, guinea pigs and
5 primates.

Antigens having superior diagnostic properties may generally be identified based on the ability to elicit a response in an intradermal skin test performed on an individual with active tuberculosis, but not in a test performed on an individual who is not infected with *M. tuberculosis*. Skin tests may generally be performed as
10 described below, with a response of at least 5 mm induration considered positive.

Immunogenic portions of the antigens described herein may be prepared and identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3d ed., Raven Press, 1993, pp. 243-247 and references cited therein. Such techniques include screening polypeptide portions of the native antigen
15 for immunogenic properties. The representative proliferation and cytokine production assays described herein may generally be employed in these screens. An immunogenic portion of a polypeptide is a portion that, within such representative assays, generates an immune response (e.g., proliferation, interferon- γ production and/or interleukin-12 production) that is substantially similar to that generated by the full length antigen. In
20 other words, an immunogenic portion of an antigen may generate at least about 20%, and preferably about 100%, of the proliferation induced by the full length antigen in the model proliferation assay described herein. An immunogenic portion may also, or alternatively, stimulate the production of at least about 20%, and preferably about 100%, of the interferon- and/or interleukin-12 induced by the full length antigen in the
25 model assay described herein.

Portions and other variants of *M. tuberculosis* antigens may be generated by synthetic or recombinant means. Synthetic polypeptides having fewer than about 100 amino acids, and generally fewer than about 50 amino acids, may be generated using techniques well known in the art.

techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. See Merrifield, *J. Am. Chem. Soc.* 85:2149-2146, 1963. Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Applied BioSystems, Inc., Foster City, CA, and may be operated according to the manufacturer's instructions. Variants of a native antigen may generally be prepared using standard mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis. Sections of the DNA sequence may also be removed using standard techniques to permit preparation of truncated polypeptides.

Recombinant polypeptides containing portions and/or variants of a native antigen may be readily prepared from a DNA sequence encoding the polypeptide using a variety of techniques well known to those of ordinary skill in the art. For example, supernatants from suitable host/vector systems which secrete recombinant protein into culture media may be first concentrated using a commercially available filter. Following concentration, the concentrate may be applied to a suitable purification matrix such as an affinity matrix or an ion exchange resin. Finally, one or more reverse phase HPLC steps can be employed to further purify a recombinant protein.

Any of a variety of expression vectors known to those of ordinary skill in the art may be employed to express recombinant polypeptides of this invention. Expression may be achieved in any appropriate host cell that has been transformed or transfected with an expression vector containing a DNA molecule that encodes a recombinant polypeptide. Suitable host cells include prokaryotes, yeast and higher eukaryotic cells. Preferably, the host cells employed are *E. coli*, yeast or a mammalian cell line such as COS or CHO. The DNA sequences expressed in this manner may encode naturally occurring antigens, portions of naturally occurring antigens, or other variants thereof.

In general, regardless of the method of preparation, the polypeptides disclosed herein are substantially pure.

preferably at least about 99% pure. In certain preferred embodiments, described in detail below, the substantially pure polypeptides are incorporated into pharmaceutical compositions or vaccines for use in one or more of the methods disclosed herein.

In certain specific embodiments, the subject invention discloses polypeptides comprising at least an immunogenic portion of a soluble *M. tuberculosis* antigen having one of the following N-terminal sequences, or a variant thereof that differs only in conservative substitutions and/or modifications:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu; (SEQ ID No. 120)
- (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser; (SEQ ID No. 121)
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg; (SEQ ID No. 122)
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro; (SEQ ID No. 123)
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val; (SEQ ID No. 124)
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro; (SEQ ID No. 125)
- (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Ala-Ala-Ala-Ser-Pro-Pro-Ser; (SEQ ID No. 126)
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly; (SEQ ID No. 127)
- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Leu-Thr-Ser-Leu-Leu-Asn-Ser-Leu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Ala-Asn; (SEQ ID No. 128)
- (j) Xaa-Asp-Ser-Glu-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID No. 129)
- (k) Xaa-Ile-Val-Thr-Ser-Lys-Ser-Ala-Thr-Ile-Lys-Val-Thr-Asp-Ala-Ser; (SEQ ID No. 130)

(f) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 136)

wherein Xaa may be any amino acid, preferably a cysteine residue. A DNA sequence encoding the antigen identified as (g) above is provided in SEQ ID No. 52, and the polypeptide encoded by SEQ ID No. 52 is provided in SEQ ID No. 53. A DNA sequence encoding the antigen defined as (a) above is provided in SEQ ID No. 101; its deduced amino acid sequence is provided in SEQ ID No. 102. A DNA sequence corresponding to antigen (d) above is provided in SEQ ID No. 24 a DNA sequence corresponding to antigen (c) is provided in SEQ ID No. 25 and a DNA sequence corresponding to antigen (i) is provided in SEQ ID No. 99; its deduced amino acid sequence is provided in SEQ ID No. 100.

In a further specific embodiment, the subject invention discloses polypeptides comprising at least an immunogenic portion of an *M. tuberculosis* antigen having one of the following N-terminal sequences, or a variant thereof that differs only in conservative substitutions and/or modifications:

(m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID No. 137) or

(n) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Xaa-Phe; (SEQ ID No. 129)

wherein Xaa may be any amino acid, preferably a cysteine residue.

In other specific embodiments, the subject invention discloses polypeptides comprising at least an immunogenic portion of a soluble *M. tuberculosis* antigen (or a variant of such an antigen) that comprises one or more of the amino acid sequences encoded by (a) the DNA sequences of SEQ ID Nos. 1, 2, 4-10, 13-25 and 52; (b) the complements of such DNA sequences, or (c) DNA sequences substantially homologous to a sequence in (a) or (b).

In further specific embodiments, the subject invention discloses polypeptides comprising at least an immunogenic portion of a *M. tuberculosis* antigen (or a variant of such an antigen) which may or may not be soluble, that comprises

Nos.: 26-51, 138, 139, 163-183, 189-193, 199, 200, 201, 203, 215-225, 239, 240, 242-247, 253-256, 261-276, 292, 293, 295-298 and 303-342. (b) the complements of such DNA sequences or (c) DNA sequences substantially homologous to a sequence in (a) or (b).

5 In the specific embodiments discussed above, the *M. tuberculosis* antigens include variants that are encoded by DNA sequences which are substantially homologous to one or more of DNA sequences specifically recited herein. "Substantial homology," as used herein, refers to DNA sequences that are capable of hybridizing under moderately stringent conditions. Suitable moderately stringent conditions include
10 prewashing in a solution of 5X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50°C-65°C, 5X SSC, overnight or, in the case of cross-species homology at 45°C, 0.5X SSC; followed by washing twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS). Such hybridizing DNA sequences are also within the scope of this invention, as are nucleotide sequences that, due to code
15 degeneracy, encode an immunogenic polypeptide that is encoded by a hybridizing DNA sequence.

In a related aspect, the present invention provides fusion proteins comprising a first and a second inventive polypeptide or, alternatively, a polypeptide of the present invention and a known *M. tuberculosis* antigen, such as the 38 kD antigen
20 described in Andersen and Hansen, *Infect. Immun.* 57:2481-2488, 1989, (Genbank Accession No. M30046) or ESAT-6 (SEQ ID Nos. 103 and 104), together with variants of such fusion proteins. The fusion proteins of the present invention may also include a linker peptide between the first and second polypeptides.

A DNA sequence encoding a fusion protein of the present invention is
25 constructed using known recombinant DNA techniques to assemble separate DNA sequences encoding the first and second polypeptides into an appropriate expression vector. The 3' end of a DNA sequence encoding the first polypeptide is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide so that the reading frames of the sequences are in phase to permit mRNA

translation of the two DNA sequences into a single fusion protein that retains the biological activity of both the first and the second polypeptides.

A peptide linker sequence may be employed to separate the first and the second polypeptides by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion protein using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea et al., *Gene* 40:39-46, 1985; Murphy et al., *Proc. Natl. Acad. Sci. USA* 83:8258-8262, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may be from 1 to about 50 amino acids in length. Peptide sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

The ligated DNA sequences are operably linked to suitable transcriptional or translational regulatory elements. The regulatory elements responsible for expression of DNA are located only 5' to the DNA sequence encoding the first polypeptides. Similarly, stop codons require to end translation and transcription termination signals are only present 3' to the DNA sequence encoding the second polypeptide.

In another aspect, the present invention provides methods for using one or more of the above polypeptides or fusion proteins (or DNA molecules encoding such polypeptides) to induce protective immunity against tuberculosis in a patient. As used herein, a "patient" refers to any mammal, including humans.

may be afflicted with a disease, or may be free of detectable disease and/or infection. In other words, protective immunity may be induced to prevent or treat tuberculosis.

In this aspect, the polypeptide, fusion protein or DNA molecule is generally present within a pharmaceutical composition and/or a vaccine. Pharmaceutical compositions may comprise one or more polypeptides, each of which may contain one or more of the above sequences (or variants thereof), and a physiologically acceptable carrier. Vaccines may comprise one or more of the above polypeptides and a non-specific immune response enhancer, such as an adjuvant or a liposome (into which the polypeptide is incorporated). Such pharmaceutical compositions and vaccines may also contain other *M. tuberculosis* antigens, either incorporated into a combination polypeptide or present within a separate polypeptide.

Alternatively, a vaccine may contain DNA encoding one or more polypeptides as described above, such that the polypeptide is generated *in situ*. In such vaccines, the DNA may be present within any of a variety of delivery systems known to those of ordinary skill in the art, including nucleic acid expression systems, bacterial and viral expression systems. Appropriate nucleic acid expression systems contain the necessary DNA sequences for expression in the patient (such as a suitable promoter and terminating signal). Bacterial delivery systems involve the administration of a bacterium (such as *Bacillus-Calmette-Guerrin*) that expresses an immunogenic portion of the polypeptide on its cell surface. In a preferred embodiment, the DNA may be introduced using a viral expression system (e.g., vaccinia or other pox virus, retrovirus, or adenovirus), which may involve the use of a non-pathogenic (defective), replication competent virus. Techniques for incorporating DNA into such expression systems are well known to those of ordinary skill in the art. The DNA may also be "naked," as described, for example, in Ulmer et al., *Science* 259:1745-1749, 1993 and reviewed by Conen, *Science* 259:1691-1692, 1993. The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells.

In a related aspect, the present invention provides a method for inducing

invention or a known *M. tuberculosis* antigen, such as the 38 kD antigen described above. For example, administration of DNA encoding a polypeptide of the present invention, either "naked" or in a delivery system as described above, may be followed by administration of an antigen in order to enhance the protective immune effect of the vaccine.

Routes and frequency of administration, as well as dosage, will vary from individual to individual and may parallel those currently being used in immunization using BCG. In general, the pharmaceutical compositions and vaccines may be administered by injection (e.g., intracutaneous, intramuscular, intravenous or subcutaneous), intranasally (e.g., by aspiration) or orally. Between 1 and 3 doses may be administered for a 1-36 week period. Preferably, 3 doses are administered, at intervals of 3-4 months, and booster vaccinations may be given periodically thereafter. Alternate protocols may be appropriate for individual patients. A suitable dose is an amount of polypeptide or DNA that, when administered as described above, is capable of raising an immune response in an immunized patient sufficient to protect the patient from *M. tuberculosis* infection for at least 1-2 years. In general, the amount of polypeptide present in a dose (or produced *in situ* by the DNA in a dose) ranges from about 1 pg to about 100 mg per kg of host, typically from about 10 pg to about 1 mg, and preferably from about 100 pg to about 1 μ g. Suitable dose sizes will vary with the size of the patient, but will typically range from about 0.1 mL to about 5 mL.

While any suitable carrier known to those of ordinary skill in the art may be employed in the pharmaceutical compositions of this invention, the type of carrier will vary depending on the mode of administration. For parenteral administration, such as subcutaneous injection, the carrier preferably comprises water, saline, alcohol, a fat, a wax or a buffer. For oral administration, any of the above carriers or a solid carrier, such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, talcum, cellulose, glucose, sucrose, and magnesium carbonate, may be employed. Biodegradable microspheres (e.g., polylactic acid/lactides) may also be employed as carriers for the pharmaceutical compositions of this invention.

Any of a variety of adjuvants may be employed in the vaccines of this invention to nonspecifically enhance the immune response. Most adjuvants contain a substance designed to protect the antigen from rapid catabolism, such as aluminum hydroxide or mineral oil, and a nonspecific stimulator of immune responses, such as lipid A, *Bordetella pertussis* or *Mycobacterium tuberculosis*. Suitable adjuvants are commercially available as, for example, Freund's Incomplete Adjuvant and Freund's Complete Adjuvant (Difco Laboratories) and Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ). Other suitable adjuvants include alum, biodegradable microspheres, monophosphoryl lipid A and quil A.

In another aspect, this invention provides methods for using one or more of the polypeptides described above to diagnose tuberculosis using a skin test. As used herein, a "skin test" is any assay performed directly on a patient in which a delayed-type hypersensitivity (DTH) reaction (such as swelling, reddening or dermatitis) is measured following intradermal injection of one or more polypeptides as described above. Such injection may be achieved using any suitable device sufficient to contact the polypeptide or polypeptides with dermal cells of the patient, such as a tuberculin syringe or 1 mL syringe. Preferably, the reaction is measured at least 48 hours after injection, more preferably 48-72 hours.

The DTH reaction is a cell-mediated immune response, which is greater in patients that have been exposed previously to the test antigen (*i.e.*, the immunogenic portion of the polypeptide employed, or a variant thereof). The response may be measured visually, using a ruler. In general, a response that is greater than about 0.5 cm in diameter, preferably greater than about 1.0 cm in diameter, is a positive response, indicative of tuberculosis infection, which may or may not be manifested as an active disease.

The polypeptides of this invention are preferably formulated, for use in a skin test, as pharmaceutical compositions containing a polypeptide and a physiologically acceptable carrier, as described above. Such compositions typically contain one or more of the above polypeptides in an amount ranging from about 0.1 mg to 10 mg.

Preferably, the carrier employed in such pharmaceutical compositions is a saline solution with appropriate preservatives, such as phenol and/or Tween 80™.

In a preferred embodiment, a polypeptide employed in a skin test is of sufficient size such that it remains at the site of injection for the duration of the reaction period. In general, a polypeptide that is at least 9 amino acids in length is sufficient. The polypeptide is also preferably broken down by macrophages within hours of injection to allow presentation to T-cells. Such polypeptides may contain repeats of one or more of the above sequences and/or other immunogenic or nonimmunogenic sequences.

The following Examples are offered by way of illustration and not by way of limitation.

EXAMPLES

EXAMPLE 1

PURIFICATION AND CHARACTERIZATION OF POLYPEPTIDES FROM *M. TUBERCULOSIS* CULTURE FILTRATE

This example illustrates the preparation of *M. tuberculosis* soluble polypeptides from culture filtrate. Unless otherwise noted, all percentages in the following example are weight per volume.

M. tuberculosis (either H37Ra, ATCC No. 25177, or H37Rv, ATCC No. 25618) was cultured in sterile GAS media at 37°C for fourteen days. The media was then vacuum filtered (leaving the bulk of the cells) through a 0.45 µ filter into a sterile 2.5 L bottle. The media was next filtered through a 0.2 µ filter into a sterile 4 L bottle and NaN₃ was added to the culture filtrate to a concentration of 0.04%. The bottles were then placed in a 4°C cold room.

The culture filtrate was concentrated by ultrafiltration.

which had been rinsed with ethanol and contained a 10,000 kDa MWCO membrane. The pressure was maintained at 60 psi using nitrogen gas. This procedure reduced the 12 L volume to approximately 50 ml.

The culture filtrate was dialyzed into 0.1% ammonium bicarbonate using
5 a 8,000 kDa MWCO cellulose ester membrane, with two changes of ammonium bicarbonate solution. Protein concentration was then determined by a commercially available BCA assay (Pierce, Rockford, IL).

The dialyzed culture filtrate was then lyophilized, and the polypeptides resuspended in distilled water. The polypeptides were dialyzed against 0.01 mM 1,3
10 bis[tris(hydroxymethyl)-methylamino]propane, pH 7.5 (Bis-Tris propane buffer), the initial conditions for anion exchange chromatography. Fractionation was performed using gel perfusion chromatography on a POROS 146 II Q/M anion exchange column 4.6 mm x 100 mm (Perseptive BioSystems, Framingham, MA) equilibrated in 0.01 mM Bis-Tris propane buffer pH 7.5. Polypeptides were eluted with a linear 0-0.5 M NaCl
15 gradient in the above buffer system. The column eluent was monitored at a wavelength of 220 nm.

The pools of polypeptides eluting from the ion exchange column were dialyzed against distilled water and lyophilized. The resulting material was dissolved in
0.1% trifluoroacetic acid (TFA) pH 1.9 in water, and the polypeptides were purified on
20 a Delta-Pak C18 column (Waters, Milford, MA) 300 Angstrom pore size, 5 micron particle size (3.9 x 150 mm). The polypeptides were eluted from the column with a linear gradient from 0-60% dilution buffer (0.1% TFA in acetonitrile). The flow rate was 0.75 ml/minute and the HPLC eluent was monitored at 214 nm. Fractions containing the eluted polypeptides were collected to maximize the purity of the
25 individual samples. Approximately 200 purified polypeptides were obtained.

The purified polypeptides were then screened for the ability to induce T-cell proliferation in PBMC preparations. The PBMCs from donors known to be PPD skin test positive and whose T-cells were shown to proliferate in response to PPD and crude soluble proteins from MTH were used for this purpose.

polypeptides were added in duplicate at concentrations of 0.5 to 10 $\mu\text{g/mL}$. After six days of culture in 96-well round-bottom plates in a volume of 200 μl , 50 μl of medium was removed from each well for determination of IFN- γ levels, as described below. The plates were then pulsed with 1 $\mu\text{Ci/well}$ of tritiated thymidine for a further 18 hours, harvested and tritium uptake determined using a gas scintillation counter. Fractions that resulted in proliferation in both replicates three fold greater than the proliferation observed in cells cultured in medium alone were considered positive.

IFN- γ was measured using an enzyme-linked immunosorbent assay (ELISA). ELISA plates were coated with a mouse monoclonal antibody directed to human IFN- γ (PharMingen, San Diego, CA) in PBS for four hours at room temperature. Wells were then blocked with PBS containing 5% (W/V) non-fat dried milk for 1 hour at room temperature. The plates were then washed six times in PBS/0.2% TWEEN-20 and samples diluted 1:2 in culture medium in the ELISA plates were incubated overnight at room temperature. The plates were again washed and a polyclonal rabbit anti-human IFN- γ serum diluted 1:3000 in PBS/10% normal goat serum was added to each well. The plates were then incubated for two hours at room temperature, washed and horseradish peroxidase-coupled anti-rabbit IgG (Sigma Chemical Co., St. Louis, MO) was added at a 1:2000 dilution in PBS/5% non-fat dried milk. After a further two hour incubation at room temperature, the plates were washed and TMB substrate added. The reaction was stopped after 20 min with 1 N sulfuric acid. Optical density was determined at 450 nm using 570 nm as a reference wavelength. Fractions that resulted in both replicates giving an OD two fold greater than the mean OD from cells cultured in medium alone, plus 3 standard deviations, were considered positive.

For sequencing, the polypeptides were individually dried onto Biobrene (Perkin Elmer/Applied BioSystems Division, Foster City, CA) treated glass fiber filters. The filters with polypeptide were loaded onto a Perkin Elmer/Applied BioSystems Division Procise 492 protein sequencer. The polypeptides were sequenced from the amino terminal and using traditional Edman chemistry. The amino acid sequence was determined for each polypeptide.

Using the procedure described above, antigens having the following N-terminal sequences were isolated:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Xaa-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu; (SEQ ID No. 54)
- (b) Ala-Val-Glu-Ser-Gly-Met-Leu-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser; (SEQ ID No. 55)
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg; (SEQ ID No. 56)
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro; (SEQ ID No. 57)
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val; (SEQ ID No. 58)
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro; (SEQ ID No. 59)
- (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Ala-Ala-Ala-Ala-Pro-Pro-Ala; (SEQ ID No. 60) and
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly; (SEQ ID No. 61)

wherein Xaa may be any amino acid.

An additional antigen was isolated employing a microbore HPLC purification step in addition to the procedure described above. Specifically, 20 μ l of a fraction comprising a mixture of antigens from the chromatographic purification step previously described, was purified on an Aquapore C18 column (Perkin Elmer Applied Biosystems Division, Foster City, CA) with a 7 micron pore size, column size 4 mm x 100 mm, in a Perkin Elmer Applied Biosystems Division Model 172 HPLC. Fractions were eluted from the column with a linear gradient of 1% minute of acetonitrile (containing 0.05% TFA) in water (0.05% TFA) at a flow rate of 80 μ l/minute. The eluent was monitored at 250 nm. The original fraction was separated into 4 major peaks plus other smaller component and was identified as follows:

have a molecular weight of 12.054 Kd (by mass spectrometry) and the following N-terminal sequence:

(i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Gln-Thr-Ser-Leu-Leu-Asn-Asn-Leu-Ala-Asp-Pro-Asp-Val-Ser-Phe-Ala-Asp (SEQ ID No. 62).

This polypeptide was shown to induce proliferation and IFN- γ production in PBMC preparations using the assays described above.

Additional soluble antigens were isolated from *M. tuberculosis* culture filtrate as follows. *M. tuberculosis* culture filtrate was prepared as described above. Following dialysis against Bis-Tris propane buffer, at pH 5.5, fractionation was performed using anion exchange chromatography on a Poros QE column 4.6 x 100 mm (Perseptive Biosystems) equilibrated in Bis-Tris propane buffer pH 5.5. Polypeptides were eluted with a linear 0-1.5 M NaCl gradient in the above buffer system at a flow rate of 10 ml/min. The column eluent was monitored at a wavelength of 214 nm.

The fractions eluting from the ion exchange column were pooled and subjected to reverse phase chromatography using a Poros R2 column 4.6 x 100 mm (Perseptive Biosystems). Polypeptides were eluted from the column with a linear gradient from 0-100% acetonitrile (0.1% TFA) at a flow rate of 5 ml/min. The eluent was monitored at 214 nm.

Fractions containing the eluted polypeptides were lyophilized and resuspended in 80 μ l of aqueous 0.1% TFA and further subjected to reverse phase chromatography on a Vydac C4 column 4.6 \times 150 mm (Western Analytical, Temecula, CA) with a linear gradient of 0-100% acetonitrile (0.1% TFA) at a flow rate of 2 ml/min. Eluent was monitored at 214 nm.

The fraction with biological activity was separated into one major peak plus other smaller components. Western blot of this peak onto PVDF membrane revealed three major bands of molecular weights 14 Kd, 20 Kd and 26 Kd. These polypeptides were determined to have the following N-terminal sequence, respectively.

(k) Ala-Gly-Asp-Thr-Xaa-Ile-Tyr-Ile-Val-Gly-Asn-Leu-Thr-Ala-Asp; (SEQ ID No. 135) and

(l) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Leu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 136), wherein Xaa may be any amino acid.

5 Using the assays described above, these polypeptides were shown to induce proliferation and IFN- γ production in PBMC preparations. Figs. 1A and B show the results of such assays using PBMC preparations from a first and a second donor, respectively.

DNA sequences that encode the antigens designated as (a), (c), (d) and
10 (g) above were obtained by screening a genomic *M. tuberculosis* library using ³²P end labeled degenerate oligonucleotides corresponding to the N-terminal sequence and containing *M. tuberculosis* codon bias. The screen performed using a probe corresponding to antigen (a) above identified a clone having the sequence provided in SEQ ID No. 101. The polypeptide encoded by SEQ ID No. 101 is provided in SEQ ID
15 No. 102. The screen performed using a probe corresponding to antigen (g) above identified a clone having the sequence provided in SEQ ID No. 52. The polypeptide encoded by SEQ ID No. 52 is provided in SEQ ID No. 53. The screen performed using a probe corresponding to antigen (d) above identified a clone having the sequence provided in SEQ ID No. 24, and the screen performed with a probe corresponding to
20 antigen (e) identified a clone having the sequence provided in SEQ ID No. 25.

The above amino acid sequences were compared to known amino acid sequences in the gene bank using the DNA STAR system. The database searched contains some 173,000 proteins and is a combination of the Swiss PIR databases along with translated protein sequences (Version 87). No significant homologies to the amino
25 acid sequences for antigens (a)-(h) and (l) were detected.

The amino acid sequence for antigen (i) was found to be homologous to a sequence from *M. leprae*. The full length *M. leprae* sequence was amplified from genomic DNA using the sequence obtained from GENBANK. This sequence was then used to screen the *M. tuberculosis* library. The clone having the sequence

The amino acid sequence for antigen (j) was found to be homologous to a known *M. tuberculosis* protein translated from a DNA sequence. To the best of the inventors' knowledge, this protein has not been previously shown to possess T-cell stimulatory activity. The amino acid sequence for antigen (k) was found to be related to a sequence from *M. leprae*.

In the proliferation and IFN- γ assays described above, using three PPD positive donors, the results for representative antigens provided above are presented in Table 1:

TABLE 1
RESULTS OF PBMC PROLIFERATION AND IFN- γ ASSAYS

Sequence	Proliferation	IFN- γ
(a)	+	-
(c)	+++	+++
(d)	++	++
(g)	+++	+++
(h)	+++	+++

In Table 1, responses that gave a stimulation index (SI) of between 2 and 4 (compared to cells cultured in medium alone) were scored as +, an SI of 4-8 or 2-4 at a concentration of 1 μ g or less was scored as ++ and an SI of greater than 8 was scored as +++. The antigen of sequence (h) was found to have a high SI (++++) for one donor and lower SI (++ and +) for the two other donors in both proliferation and IFN- γ assays. These results indicate that these antigens are capable of inducing proliferation and/or interferon- γ production.

EXAMPLE 2

USE OF PATIENT SERA TO ISOLATE *M. TUBERCULOSIS* ANTIGENS

This example illustrates the isolation of antigens from *M. tuberculosis* lysate by screening with serum from *M. tuberculosis*-infected individuals.

Dessicated *M. tuberculosis* H37Ra (Difco Laboratories) was added to a 2% NP40 solution, and alternately homogenized and sonicated three times. The resulting suspension was centrifuged at 13,000 rpm in microfuge tubes and the supernatant put through a 0.2 micron syringe filter. The filtrate was bound to Macro Prep DEAE beads (BioRad, Hercules, CA). The beads were extensively washed with 20 mM Tris pH 7.5 and bound proteins eluted with 1M NaCl. The 1M NaCl elute was dialyzed overnight against 10 mM Tris, pH 7.5. Dialyzed solution was treated with DNase and RNase at 0.05 mg/ml for 30 min. at room temperature and then with α -D-mannosidase, 0.5 U/mg at pH 4.5 for 3-4 hours at room temperature. After returning to pH 7.5, the material was fractionated via FPLC over a Bio Scale-Q-20 column (BioRad). Fractions were combined into nine pools, concentrated in a Centriprep 10 (Amicon, Beverley, MA) and then screened by Western blot for serological activity using a serum pool from *M. tuberculosis*-infected patients which was not immunoreactive with other antigens of the present invention.

The most reactive fraction was run in SDS-PAGE and transferred to PVDF. A band at approximately 85 Kd was cut out yielding the sequence:

(m) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val SEQ ID No. 137, wherein Xaa may be any amino acid

Comparison of this sequence with those in the gene bank as described above, revealed no significant homologies to known sequences.

A DNA sequence that encodes the antigen designated as (m) above was obtained by screening a genomic *M. tuberculosis* Erdman strain library using labeled degenerate oligonucleotides corresponding to the amino acid sequence (m).

203. This sequence was found to encode the amino acid sequence provided in SEQ ID
NO: 204. Comparison of these sequences with those in the genebank revealed some
similarity to sequences previously identified in *M. tuberculosis* and *M. bovis*.

5

EXAMPLE 3

PREPARATION OF DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

This example illustrates the preparation of DNA sequences encoding
M. tuberculosis antigens by screening a *M. tuberculosis* expression library with sera
10 obtained from patients infected with *M. tuberculosis*, or with anti-sera raised against
soluble *M. tuberculosis* antigens.

A. PREPARATION OF *M. TUBERCULOSIS* SOLUBLE ANTIGENS USING RABBIT ANTI-SERA RAISED AGAINST *M. TUBERCULOSIS* SUPERNATANT

15 Genomic DNA was isolated from the *M. tuberculosis* strain H37Ra. The
DNA was randomly sheared and used to construct an expression library using the
Lambda ZAP expression system (Stratagene, La Jolla, CA). Rabbit anti-sera was
generated against secretory proteins of the *M. tuberculosis* strains H37Ra, H37Rv and
Erđman by immunizing a rabbit with concentrated supernatant of the *M. tuberculosis*
20 cultures. Specifically, the rabbit was first immunized subcutaneously with 200 µg of
protein antigen in a total volume of 2 ml containing 10 µg muramyl dipeptide
(Calbiochem, La Jolla, CA) and 1 ml of incomplete Freund's adjuvant. Four weeks later
the rabbit was boosted subcutaneously with 100 µg antigen in incomplete Freund's
adjuvant. Finally, the rabbit was immunized intravenously four weeks later with 50 µg
25 protein antigen. The anti-sera were used to screen the expression library as described in
Samorook et al., *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor
Laboratories, Cold Spring Harbor, NY, 1989. Bacteriophage plaques expressing
immunoreactive antigens were purified. Phagemid from the plaques was rescued and
the nucleotide sequences of the *M. tuberculosis* clones deduced.

were expressed and purified antigens used in the immunological analysis described in Example 1. Proteins were induced by IPTG and purified by gel elution, as described in Skeiky et al., *J. Exp. Med.* 181:1527-1537, 1995. Representative sequences of DNA molecules identified in this screen are provided in SEQ ID Nos. 1-25. The
5 corresponding predicted amino acid sequences are shown in SEQ ID Nos. 63-87.

On comparison of these sequences with known sequences in the gene bank using the databases described above, it was found that the clones referred to hereinafter as TbRA2A, TbRA16, TbRA18, and TbRA29 (SEQ ID Nos. 76, 68, 70, 75) show some homology to sequences previously identified in *Mycobacterium leprae* but
10 not in *M. tuberculosis*. TbRA2A was found to be a lipoprotein, with a six residue lipidation sequence being located adjacent to a hydrophobic secretory sequence. TbRA11, TbRA26, TbRA28 and TbDPEP (SEQ ID Nos. 65, 73, 74, 53) have been previously identified in *M. tuberculosis*. No significant homologies were found to
15 TbRA1, TbRA3, TbRA4, TbRA9, TbRA10, TbRA13, TbRA17, TbRa19, TbRA29, TbRA32, TbRA36 and the overlapping clones TbRA35 and TbRA12 (SEQ ID Nos. 63, 77, 81, 82, 64, 67, 69, 71, 75, 78, 80, 79, 66). The clone TbRa24 is overlapping with clone TbRa29.

The results of PBMC proliferation and interferon- γ assays performed on representative recombinant antigens, and using T-cell preparations from several
20 different *M. tuberculosis*-immune patients, are presented in Tables 2 and 3, respectively.

TABLE 2
RESULTS OF PBMC PROLIFERATION TO REPRESENTATIVE SOLUBLE ANTIGENS

Antigen	Patient												
	1	2	3	4	5	6	7	8	9	10	11	12	13
IbRa1	-	-	+	+	-	-	+	+	-	-	+	+	-
IbRa3	-	-	+	-	+	-	-	++	+	-	-	+	-
IbRa9	-	-	nt	nt	++	+	nt	nt	nt	nt	nt	nt	nt
IbRa10	-	-	+	+	+	+	nt	+	-	+	+	+	+
IbRa11	-	-	+	+	+	+	nt	+	-	+	+	+	-
IbRa12	-	-	+	+	+	+	nt	+	+	+	+	+	nt
IbRa16	-	-	nt	nt	-	+	nt	nt	nt	nt	nt	nt	-
IbRa24	-	-	nt	nt	-	-	nt	nt	nt	nt	nt	nt	nt
IbRa26	-	-	nt	nt	-	-	nt	nt	nt	nt	nt	nt	nt
IbRa29	-	-	nt	nt	-	-	nt	nt	nt	nt	nt	nt	nt
IbRa35	-	-	nt	nt	++	++	nt	++	++	++	++	++	nt
IbRaB	-	-	nt	nt	-	-	nt	nt	nt	nt	nt	nt	nt
IbRaC	-	-	nt	nt	-	-	nt	nt	nt	nt	nt	nt	nt
IbRaD	-	-	nt	nt	-	-	nt	nt	nt	nt	nt	nt	nt
AAMK	-	-	+	-	-	-	nt	nt	nt	nt	nt	nt	nt
YY	-	-	-	-	-	-	nt	-	-	-	nt	+	nt
DPEP	-	-	-	-	-	-	nt	++	+	+	nt	+	nt
Control	-	-	-	-	-	-	-	-	-	-	-	-	-

nt tested

In Tables 2 and 3, responses that gave a stimulation index (SI) of between 1.2 and 2 (compared to cells cultured in medium alone) were scored as \pm , a SI of 2-4 was scored as +, as SI of 4-8 or 2-4 at a concentration of 1 μ g or less was scored as ++ and an SI of greater than 8 was scored as ++++. In addition, the effect of concentration on proliferation and interferon- γ production is shown for two of the above antigens in the attached Figure. For both proliferation and interferon- γ production, TbRa3 was scored as ++ and TbRa9 as +.

These results indicate that these soluble antigens can induce proliferation and/or interferon- γ production in T-cells derived from an *M. tuberculosis*-immune individual.

B. USE OF SERA FROM PATIENTS HAVING PULMONARY OR PLEURAL TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

The genomic DNA library described above, and an additional H37Rv library, were screened using pools of sera obtained from patients with active tuberculosis. To prepare the H37Rv library, *M. tuberculosis* strain H37Rv genomic DNA was isolated, subjected to partial *Sau3A* digestion and used to construct an expression library using the Lambda Zap expression system (Stratagene, La Jolla, Ca). Three different pools of sera, each containing sera obtained from three individuals with active pulmonary or pleural disease, were used in the expression screening. The pools were designated TbL, TbM and TbH, referring to relative reactivity with H37Ra lysate (i.e., TbL = low reactivity, TbM = medium reactivity and TbH = high reactivity) in both ELISA and immunoblot format. A fourth pool of sera from seven patients with active pulmonary tuberculosis was also employed. All of the sera lacked increased reactivity with the recombinant 38 kD *M. tuberculosis* H37Ra phosphate-binding protein.

All pools were pre-adsorbed with *E. coli* lysate and used to screen the H37Ra and H37Rv expression libraries, as described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989. Bacteriophage plaques expressing immunoreactive antigens were purified

Thirty two clones were purified. Of these, 31 represented sequences that had not been previously identified in human *M. tuberculosis*. Representative sequences of the DNA molecules identified are provided in SEQ ID Nos.: 26-51 and 105. Of these, TbH-8-2 (SEQ. ID NO. 105) is a partial clone of TbH-8, and TbH-4 (SEQ. ID NO. 43) and TbH-4-FWD (SEQ. ID NO. 44) are non-contiguous sequences from the same clone. Amino acid sequences for the antigens hereinafter identified as Tb38-1, TbH-4, TbH-8, TbH-9, and TbH-12 are shown in SEQ ID Nos.: 88-92. Comparison of these sequences with known sequences in the gene bank using the databases identified above revealed no significant homologies to TbH-4, TbH-8, TbH-9 and TbM-3, although weak homologies were found to TbH-9. TbH-12 was found to be homologous to a 34 kD antigenic protein previously identified in *M. paratuberculosis* (Acc. No. S28515). Tb38-1 was found to be located 34 base pairs upstream of the open reading frame for the antigen ESAT-6 previously identified in *M. bovis* (Acc. No. U34848) and in *M. tuberculosis* (Sorensen et al., *Infect. Immun.* 63:1710-1717, 1995).


Probes derived from Tb38-1 and TbH-9, both isolated from an H37Ra library, were used to identify clones in an H37Rv library. Tb38-1 hybridized to Tb38-1F2, Tb38-1F3, Tb38-1F5 and Tb38-1F6 (SEQ. ID NOS. 112, 113, 116, 118, and 119). (SEQ ID NOS. 112 and 113 are non-contiguous sequences from clone Tb38-1F2). Two open reading frames were deduced in Tb38-1F2: one corresponds to Tb37FL (SEQ. ID NO. 114), the second, a partial sequence, may be the homologue of Tb38-1 and is called Tb38-IN (SEQ. ID NO. 115). The deduced amino acid sequence of Tb38-1F3 is presented in SEQ. ID NO. 117. A TbH-9 probe identified three clones in the H37Rv library: TbH-9-FL (SEQ. ID NO. 106), which may be the homologue of TbH-9 (H37Ra), TbH-9-1 (SEQ. ID NO. 108), and TbH-9-4 (SEQ. ID NO. 110), all of which are highly related sequences to TbH-9. The deduced amino acid sequences for these three clones are presented in SEQ ID NOS. 107, 109 and 111.

Further screening of the *M. tuberculosis* genomic DNA library, as described above, resulted in the identification of additional clones.

above, one was determined to be identical to the 14Kd alpha crystallin heat shock protein previously shown to be present in *M. tuberculosis*, and a third was determined to be identical to the antigen TbH-8 described above. The determined DNA sequences for the remaining five clones (hereinafter referred to as TbH-29, TbH-30, TbH-32 and 5 TbH-33) are provided in SEQ ID NO: 138-141, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 142-145, respectively. The DNA and amino acid sequences for these antigens were compared with those in the gene bank as described above. No homologies were found to the 5' end of TbH-29, 10 (which contains the reactive open reading frame), although the 3' end of TbH-29 was found to be identical to the *M. tuberculosis* cosmid Y227. TbH-32 and TbH-33 were found to be identical to the previously identified *M. tuberculosis* insertion element IS6110 and to the *M. tuberculosis* cosmid Y50, respectively. No significant homologies to TbH-30 were found.

Positive phagemid from this additional screening were used to infect *E.* 15 *coli* XL-1 Blue MRF', as described in Sambrook et al., *supra*. Induction of recombinant protein was accomplished by the addition of IPTG. Induced and uninduced lysates were run in duplicate on SDS-PAGE and transferred to nitrocellulose filters. Filters were reacted with human *M. tuberculosis* sera (1:200 dilution) reactive with TbH and a rabbit sera (1:200 or 1:250 dilution) reactive with the N-terminal 4 Kd portion of lacZ. 20 Sera incubations were performed for 2 hours at room temperature. Bound antibody was detected by addition of ¹²⁵I-labeled Protein A and subsequent exposure to film for variable times ranging from 16 hours to 11 days. The results of the immunoblots are summarized in Table 4

TABLE 4

Antigen	Human <i>M. tb</i> Sera	Anti-lacZ Sera
TbH-29	45 Kd	45 Kd
TbH-30	No reactivity	29 Kd
TbH-32	12 Kd	12 Kd
TbH-33	16 Kd 	16 Kd

10

Positive reaction of the recombinant human *M. tuberculosis* antigens with both the human *M. tuberculosis* sera and anti-lacZ sera indicate that reactivity of the human *M. tuberculosis* sera is directed towards the fusion protein. Antigens reactive with the anti-lacZ sera but not with the human *M. tuberculosis* sera may be the result of the human *M. tuberculosis* sera recognizing conformational epitopes, or the antigen-antibody binding kinetics may be such that the 2 hour sera exposure in the immunoblot is not sufficient.

The results of T-cell assays performed on Tb38-1, ESAT-6 and other representative recombinant antigens are presented in Tables 5A, B and C, respectively, below:

TABLE 5A
RESULTS OF PBMC PROLIFERATION TO REPRESENTATIVE ANTIGENS

Antigen	Donor										
	1	2	3	4	5	6	7	8	9	10	11
Tb38-1	—	+	+	+	+	+	+	+	+	+	+
ESAT-6	—	+	+	+	+	+	+	+	+	+	+
TbH-29	—	—	—	—	—	—	—	—	—	—	—

15

TABLE 5B
RESULTS OF PBMC INTERFERON- γ PRODUCTION TO REPRESENTATIVE ANTIGENS

Antigen	Donor										
	1	2	3	4	5	6	7	8	9	10	11
Tb38.1	+++	+	-	+	+	---	-	++	-	---	---
ESAT-6	+++	+	-	+	+-	-	-	+	-	---	---
TbH-9	++	++	-	+++	±	±	---	---	++	---	++

5

TABLE 6
SUMMARY OF T-CELL RESPONSES TO REPRESENTATIVE ANTIGENS

Antigen	Proliferation			Interferon- γ			total
	patient 4	patient 5	patient 6	patient 4	patient 5	patient 6	
TbH9	++	++	++	---	++	++	13
TbM7	-	-	-	++	+	-	4
TbH5	-	+	+	++	++	++	8
TbL23	-	+	=	++	---	+	7.5
TbH4	-	---	=	---	---	±	7
- control	-	-	-	-	-	-	0

10

These results indicate that both the inventive *M. tuberculosis* antigens and ESAT-6 can induce proliferation and/or interferon- γ production in T-cells derived from an *M. tuberculosis*-immune individual. To the best of the inventors' knowledge, ESAT-6 has not been previously shown to stimulate human immune responses

A set of six overlapping peptides covering the amino acid sequence of the antigen Tb38-1 was constructed using the method described in Example 6. The sequences of these peptides, hereinafter referred to as pep1-6, are provided in SEQ ID Nos. 93-98, respectively. The results of T-cell assays using these peptides are shown in Tables 7 and 8. These results confirm the existence, and help to localize T-cell epitopes within Tb38-1 capable of inducing proliferation and interferon- γ production in T-cells.

15

TABLE 7
RESULTS OF PBMC PROLIFERATION TO TB38.1 PEPTIDES

[illegible]

Studies were undertaken to determine whether the antigens TbH-9 and Tb38-1 represent cellular proteins or are secreted into *M. tuberculosis* culture media. In the first study, rabbit sera were raised against A) secretory proteins of *M. tuberculosis*, B) the known secretory recombinant *M. tuberculosis* antigen 85b, C) recombinant Tb38-1 and D) recombinant TbH-9, using protocols substantially the same as that as described in Example 3A. Total *M. tuberculosis* lysate, concentrated supernatant of *M. tuberculosis* cultures and the recombinant antigens 85b, TbH-9 and Tb38-1 were resolved on denaturing gels, immobilized on nitrocellulose membranes and duplicate blots were probed using the rabbit sera described above.

The results of this analysis using control sera (panel I) and antisera (panel II) against secretory proteins, recombinant 85b, recombinant Tb38-1 and recombinant TbH-9 are shown in Figures 3A-D, respectively, wherein the lane designations are as follows: 1) molecular weight protein standards; 2) 5 µg of *M. tuberculosis* lysate; 3) 5 µg secretory proteins; 4) 50 ng recombinant Tb38-1; 5) 50 ng recombinant TbH-9; and 6) 50 ng recombinant 85b. The recombinant antigens were engineered with six terminal histidine residues and would therefore be expected to migrate with a mobility approximately 1 kD larger than the native protein. In Figure 3D, recombinant TbH-9 is lacking approximately 10 kD of the full-length 42 kD antigen, hence the significant difference in the size of the immunoreactive native TbH-9 antigen in the lysate lane (indicated by an arrow). These results demonstrate that Tb38-1 and TbH-9 are intracellular antigens and are not actively secreted by *M. tuberculosis*.

The finding that TbH-9 is an intracellular antigen was confirmed by determining the reactivity of TbH-9-specific human T cell clones to recombinant TbH-9, secretory *M. tuberculosis* proteins and PPD. A TbH-9-specific T cell clone (designated 131TbH-9) was generated from PBMC of a healthy PPD-positive donor. The proliferative response of 131TbH-9 to secretory proteins, recombinant TbH-9 and a control *M. tuberculosis* antigen, TbRa11, was determined by measuring uptake of tritiated thymidine, as described in Example 1. As shown in Figure 4A, the clone 131TbH-9 responds specifically to TbH-9, showing that TbH-9 is not a significant component of the secretory products.

(designated PPD 800-10) prepared from PBMC from a healthy PPD-positive donor, following stimulation of the T cell clone with secretory proteins, PPD or recombinant TbH-9. These results further confirm that TbH-9 is not secreted by *M. tuberculosis*.

C. USE OF SERA FROM PATIENTS HAVING EXTRAPULMONARY TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

Genomic DNA was isolated from *M. tuberculosis* Erdman strain, randomly sheared and used to construct an expression library employing the Lambda ZAP expression system (Stratagene, La Jolla, CA). The resulting library was screened using pools of sera obtained from individuals with extrapulmonary tuberculosis, as described above in Example 3B, with the secondary antibody being goat anti-human IgG + A + M (H+L) conjugated with alkaline phosphatase.

Eighteen clones were purified. Of these, 4 clones (hereinafter referred to as XP14, XP24, XP31 and XP32) were found to bear some similarity to known sequences. The determined DNA sequences for XP14, XP24 and XP31 are provided in SEQ ID Nos. 156-158, respectively, with the 5' and 3' DNA sequences for XP32 being provided in SEQ ID Nos. 159 and 160, respectively. The predicted amino acid sequence for XP14 is provided in SEQ ID No. 161. The reverse complement of XP14 was found to encode the amino acid sequence provided in SEQ ID No. 162.

Comparison of the sequences for the remaining 14 clones (hereinafter referred to as XP1-XP6, XP17-XP19, XP22, XP25, XP27, XP30 and XP36) with those in the genbank as described above, revealed no homologies with the exception of the 3' ends of XP2 and XP6 which were found to bear some homology to known *M. tuberculosis* cosmid. The DNA sequences for XP27 and XP36 are shown in SEQ ID Nos. 163 and 164, respectively, with the 5' sequences for XP4, XP5, XP17 and XP30 being shown in SEQ ID Nos. 165-168, respectively, and the 5' and 3' sequences for XP2, XP3, XP6, XP18, XP19, XP22 and XP25 being shown in SEQ ID Nos. 169 and 170, 171 and 172, 173 and 174, 175 and 176, 177 and 178, 179 and 180, and 181 and 182, respectively. XP1 was found to

open reading frame encoding the amino acid sequence shown in SEQ ID No: 184. The reverse complement of TbH4-XP1 was found to contain an open reading frame encoding the amino acid sequence shown in SEQ ID No.: 185. The DNA sequence for XP36 was found to contain two open reading frames encoding the amino acid sequence shown in SEQ ID Nos.: 186 and 187, with the reverse complement containing an open reading frame encoding the amino acid sequence shown in SEQ ID No.: 188.

Recombinant XP1 protein was prepared as described above in Example 3B, with a metal ion affinity chromatography column being employed for purification. As illustrated in Figures 8A-B and 9A-B, using the assays described herein, recombinant XP1 was found to stimulate cell proliferation and IFN- γ production in T cells isolated from an *M. tuberculosis*-immune donors.

D. USE OF A LYSATE POSITIVE SERUM POOL FROM PATIENTS HAVING TUBERCULOSIS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

Genomic DNA was isolated from *M. tuberculosis* Erdman strain, randomly sheared and used to construct an expression library employing the Lambda Screen expression system (Novagen, Madison, WI), as described below in Example 6. Pooled serum obtained from *M. tuberculosis*-infected patients and that was shown to react with *M. tuberculosis* lysate but not with the previously expressed proteins 38kD, Tb38-1, TbRa3, TbH4, DPEP and TbRa11, was used to screen the expression library as described above in Example 3B, with the secondary antibody being goat anti-human IgG - A - M (H+L) conjugated with alkaline phosphatase.

Twenty-seven clones were purified. Comparison of the determined cDNA sequences for these clones revealed no significant homologies to 10 of the clones (hereinafter referred to as LSER-10, LSER-11, LSER-12, LSER-13, LSER-16, LSER-18, LSER-23, LSER-24, LSER-25 and LSER-27). The determined cDNA sequences for LSER-10, LSER-11, LSER-12, LSER-13, LSER-16 and LSER-25 are provided in SEQ ID Nos. 242-247, respectively, with the corresponding predicted amino acid sequences for LSER-10, LSER-11, LSER-12, LSER-13, LSER-16 and LSER-25 provided in SEQ ID Nos. 248-253, respectively.

and LSER-27 are shown in SEQ ID NO: 253-256, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 257-260. The remaining seventeen clones were found to show similarities to unknown sequences previously identified in *M. tuberculosis*. The determined 5' cDNA sequences for sixteen of these clones (hereinafter referred to as LSER-1, LSER-3, LSER-4, LSER-5, LSER-6, LSER-8, LSER-14, LSER-15, LSER-17, LSER-19, LSER-20, LSER-22, LSER-26, LSER-28, LSER-29 and LSER-30) are provided in SEQ ID NO: 261-276, respectively, with the corresponding predicted amino acid sequences for LSER-1, LSER-3, LSER-5, LSER-6, LSER-8, LSER-14, LSER-15, LSER-17, LSER-19, LSER-20, LSER-22, LSER-26, LSER-28, LSER-29 and LSER-30 being provided in SEQ ID NO: 277-291, respectively. The determined full-length cDNA sequence for the clone LSER-9 is provided in SEQ ID NO: 292. The reverse complement of LSER-6 (SEQ ID NO: 293) was found to encode the predicted amino acid sequence of SEQ ID NO: 294.

E. PREPARATION OF *M. TUBERCULOSIS* SOLUBLE ANTIGENS USING RABBIT ANTI-SERA RAISED AGAINST *M. TUBERCULOSIS* FRACTIONATED PROTEINS

M. tuberculosis lysate was prepared as described above in Example 2. The resulting material was fractionated by HPLC and the fractions screened by Western blot for serological activity with a serum pool from *M. tuberculosis*-infected patients which showed little or no immunoreactivity with other antigens of the present invention. Rabbit anti-sera was generated against the most reactive fraction using the method described in Example 3A. The anti-sera was used to screen an *M. tuberculosis* Erdman strain genomic DNA expression library prepared as described above. Bacteriophage plaques expressing immunoreactive antigens were purified. Phagemid from the plaques was rescued and the nucleotide sequences of the *M. tuberculosis* clones determined.

Ten different clones were purified. Of these, one was found to be TbRa35, described above, and one was found to be the previously identified *M. tuberculosis* antigen, HSP60. Of the remaining eight clones, seven hereinafter referred to as RDIF2, RDIF5, RDIF8, RDIF10, RDIF11 and RDIF12 were found to bear some similarity to previously

corresponding predicted amino acid sequences being provided in SEQ ID Nos: 194-198, respectively. The 5' and 3' DNA sequences for RDIF12 are provided in SEQ ID Nos.: 199 and 200, respectively. No significant homologies were found to the antigen RDIF-7. The determined DNA and predicted amino acid sequences for RDIF7 are provided in SEQ ID Nos.: 201 and 202, respectively. One additional clone, referred to as RDIF6 was isolated, however, this was found to be identical to RDIF5.

Recombinant RDIF6, RDIF8, RDIF10 and RDIF11 were prepared as described above. As shown in Figures 8A-B and 9A-B, these antigens were found to stimulate cell proliferation and IFN- γ production in T cells isolated from *M. tuberculosis*-immune donors.

EXAMPLE 4

PURIFICATION AND CHARACTERIZATION OF A POLYPEPTIDE FROM TUBERCULIN PURIFIED PROTEIN DERIVATIVE

An *M. tuberculosis* polypeptide was isolated from tuberculin purified protein derivative (PPD) as follows.

PPD was prepared as published with some modification (Seibert, F. et al., Tuberculin purified protein derivative. Preparation and analyses of a large quantity for standard. The American Review of Tuberculosis 44:9-25, 1941).

M. tuberculosis Rv strain was grown for 6 weeks in synthetic medium in roller bottles at 37°C. Bottles containing the bacterial growth were then heated to 100° C in water vapor for 3 hours. Cultures were sterile filtered using a 0.22 μ filter and the liquid phase was concentrated 20 times using a 3 kD cut-off membrane. Proteins were precipitated once with 50% ammonium sulfate solution and eight times with 25% ammonium sulfate solution. The resulting proteins (PPD) were fractionated by reverse phase liquid chromatography (RP HPLC) using a C18 column (7.8 x 300 mM; Waters, Milford, MA) in a Biocad HPLC system (Perseptive Biosystems, Framingham, MA). Fractions were eluted from the column with a linear gradient from 0-100% buffer (0.1% TFA in acetonitrile). The flow rate was 1.0 ml/minute and eluent was monitored at 214 nm and 280 nm.

Six fractions were collected, dried, suspended in PBS and tested individually in *M. tuberculosis*-infected guinea pigs for induction of delayed type hypersensitivity (DTH) reaction. One fraction was found to induce a strong DTH reaction and was subsequently fractionated further by RP-HPLC on a microbore Vydac C18 column (Cat. No. 218TP5115) in a Perkin Elmer/Applied Biosystems Division Model 172 HPLC. Fractions were eluted with a linear gradient from 5-100% buffer (0.05% TFA in acetonitrile) with a flow rate of 80 μ l/minute. Eluent was monitored at 215 nm. Eight fractions were collected and tested for induction of DTH in *M. tuberculosis*-infected guinea pigs. One fraction was found to induce strong DTH of about 16 mm induration. The other fractions did not induce detectable DTH. The positive fraction was submitted to SDS-PAGE gel electrophoresis and found to contain a single protein band of approximately 12 kD molecular weight.

This polypeptide, herein after referred to as DPPD, was sequenced from the amino terminal using a Perkin Elmer/Applied Biosystems Division Procise 492 protein sequencer as described above and found to have the N-terminal sequence shown in SEQ ID No.: 129. Comparison of this sequence with known sequences in the gene bank as described above revealed no known homologies. Four cyanogen bromide fragments of DPPD were isolated and found to have the sequences shown in SEQ ID Nos.: 130-133. A subsequent search of the *M. tuberculosis* genome database released by the Institute for Genomic Research revealed a match of the DPPD partial amino acid sequence with a sequence present within the *M. tuberculosis* cosmid MTY21C12. An open reading frame of 336 bp was identified. The full-length DNA sequence for DPPD is provided in SEQ ID NO: 240, with the corresponding full-length amino acid sequence being provided in SEQ ID NO: 241.

The ability of the antigen DPPD to stimulate human PBMC to proliferate and to produce IFN- γ was assayed as described in Example 1. As shown in Table 9, DPPD was found to stimulate proliferation and elicit production of large quantities of IFN- γ , more than that elicited by commercial PPD.

TABLE 9
RESULTS OF PROLIFERATION AND INTERFERON- γ ASSAYS TO DPPD

PBMC Donor	Stimulator	Proliferation (CPM)	IFN- γ (OD ₄₅₀)
A	Medium	1,089	0.17
	PPD (commercial)	8,394	1.29
	DPPD	13,451	2.21
B	Medium	450	0.09
	PPD (commercial)	3,929	1.26
	DPPD	6,184	1.49
C	Medium	541	0.11
	PPD (commercial)	8,907	0.76
	DPPD	23,024	>2.70

EXAMPLE 5

USE OF SERA FROM TUBERCULOSIS-INFECTED MONKEYS TO IDENTIFY DNA SEQUENCES ENCODING *M. TUBERCULOSIS* ANTIGENS

Genomic DNA was isolated from *M. tuberculosis* Erdman strain, randomly sheared and used to construct an expression library employing the Lambda ZAP expression system (Stratagene, La Jolla, CA). Serum samples were obtained from a cynomolgous monkey 18, 33, 51 and 56 days following infection with *M. tuberculosis* Erdman strain. These samples were pooled and used to screen the *M. tuberculosis* genomic DNA expression library using the procedure described above in Example 3C.

Twenty clones were purified. The determined 5' DNA sequences for the clones referred to as MO-1, MO-2, MO-4, MO-8, MO-9, MO-26, MO-28, MO-29, MO-30, MO-34 and MO-35 are provided in SEQ ID NO: 213-225, respectively, with the corresponding

predicted amino acid sequence being provided in SEQ ID NO: 238. The 3' DNA sequence for the clone MO-27 is provided in SEQ ID NO: 239.

Clones MO-1, MO-30 and MO-35 were found to show a high degree of relatedness and showed some homology to a previously identified unknown *M. tuberculosis* sequence and to cosmid MTC1237. MO-2 was found to show some homology to aspartokinase from *M. tuberculosis*. Clones MO-3, MO-7 and MO-27 were found to be identical and to show a high degree of relatedness to MO-5. All four of these clones showed some homology to *M. tuberculosis* heat shock protein 70. MO-27 was found to show some homology to *M. tuberculosis* cosmid MTCY339. MO-4 and MO-34 were found to show some homology to cosmid SCY21B4 and *M. smegmatis* integration host factor, and were both found to show some homology to a previously identified, unknown *M. tuberculosis* sequence. MO-6 was found to show some homology to *M. tuberculosis* heat shock protein 65. MO-8, MO-9, MO-10, MO-26 and MO-29 were found to be highly related to each other and to show some homology to *M. tuberculosis* dihydrolipamide succinyltransferase. MO-28, MO-31 and MO-32 were found to be identical and to show some homology to a previously identified *M. tuberculosis* protein. MO-33 was found to show some homology to a previously identified 14 kDa *M. tuberculosis* heat shock protein.

Further studies using the above protocol resulted in the isolation of an additional four clones, hereinafter referred to as MO-12, MO-13, MO-19 and MO-39. The determined 5' cDNA sequences for these clones are provided in SEQ ID NO: 295-298, respectively, with the corresponding predicted protein sequences being provided in SEQ ID NO: 299-302, respectively. Comparison of these sequences with those in the gene bank as described above revealed no significant homologies to MO-39. MO-12, MO-13 and MO-19 were found to show some homologies to unknown sequences previously isolated from *M. tuberculosis*.

EXAMPLE 6

ISOLATION OF DNA SEQUENCES ENCODING *M. tuberculosis* ANTIGENS

This example illustrates isolation of DNA sequences encoding *M. tuberculosis* antigens by screening of a novel expression library with sera from *M. tuberculosis*-infected patients that were shown to be unreactive with a panel of the recombinant *M. tuberculosis* antigens TbRa11, TbRa3, Tb38-1, TbH4, TbF and 38 kD.

Genomic DNA from *M. tuberculosis* Erdman strain was randomly sheared to an average size of 2 kb, and blunt ended with Klenow polymerase, followed by the addition of EcoRI adaptors. The insert was subsequently ligated into the Screen phage vector (Novagen, Madison, WI) and packaged *in vitro* using the PhageMaker extract (Novagen). The resulting library was screened with sera from several *M. tuberculosis* donors that had been shown to be negative on a panel of previously identified *M. tuberculosis* antigens as described above in Example 3B.

A total of 22 different clones were isolated. By comparison, screening of the λ Zap library described above using the same sera did not result in any positive hits. One of the clones was found to represent TbRa11, described above. The determined 5' cDNA sequences for 19 of the remaining 21 clones (hereinafter referred to as Erdsn1, Erdsn2, Erdsn4-Erdsn10, Erdsn12-18, Erdsn21-Erdsn23 and Erdsn25) are provided in SEQ ID NO: 303-322, respectively, with the determined 3' cDNA sequences for Erdsn1, Erdsn2, Erdsn4, Erdsn-5, Erdsn-7-Erdsn10, Erdsn12-18, Erdsn21-Erdsn23 and Erdsn25 being provided in SEQ ID NO: 323-341, respectively. The complete cDNA insert sequence for the clone Erdsn24 is provided in SEQ ID NO: 342. Comparison of the determined cDNA sequences with those in the gene bank revealed no significant homologies to the sequences provided in SEQ ID NO: 309, 316, 318-320, 322, 324, 328, 329, 333, 335, 337, 339 and 341. The sequences of SEQ ID NO: 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342 were found to show some homology to unknown sequences previously identified in *M. tuberculosis*.

EXAMPLE 7

ISOLATION OF SOLUBLE MATERIALS FROM POLYMERIZATION MIXTURES

This example illustrates the use of mass spectrometry to identify soluble *M. tuberculosis* antigens.

In a first approach, *M. tuberculosis* culture filtrate was screened by Western analysis using serum from a tuberculosis-infected individual. The reactive bands were excised from a silver stained gel and the amino acid sequences determined by mass spectrometry. The determined amino acid sequence for one of the isolated antigens is provided in SEQ ID NO: 343. Comparison of this sequence with those in the gene bank revealed homology to the 85b precursor antigen previously identified in *M. tuberculosis*.

In a second approach, the high molecular weight region of *M. tuberculosis* culture supernatant was studied. This area may contain immunodominant antigens which may be useful in the diagnosis of *M. tuberculosis* infection. Two known monoclonal antibodies, IT42 and IT57 (available from the Center for Disease Control, Atlanta, GA), show reactivity by Western analysis to antigens in this vicinity, although the identity of the antigens remains unknown. In addition, unknown high-molecular weight proteins have been described as containing a surrogate marker for *M. tuberculosis* infection in HIV-positive individuals (*Jnl. Infect. Dis.*, 176:133-143, 1997). To determine the identity of these antigens, two-dimensional gel electrophoresis and two-dimensional Western analysis were performed using the antibodies IT57 and IT42. Five protein spots in the high molecular weight region were identified, individually excised, enzymatically digested and subjected to mass spectrometric analysis.

The determined amino acid sequences for three of these spots (referred to as spots 1, 2 and 4) are provided in SEQ ID NO: 344, 345-346 and 347, respectively. Comparison of these sequences with those in the gene bank revealed that spot 1 is the previously identified Pck-1, a phosphoenolpyruvate kinase. The two sequences isolated from spot 2 were determined to be from two DNAs, previously identified in *M. tuberculosis* as heat shock proteins. Spot 4 was determined to be the previously identified *M. tuberculosis* protein Kat G. To the best of the inventors' knowledge, neither Pck-1 nor the two DNAs have previously been shown to have utility in the diagnosis of *M. tuberculosis* infection.

USE OF REPRESENTATIVE ANTIGENS FOR DIAGNOSIS OF TUBERCULOSIS

This example illustrates the effectiveness of several representative polypeptides in skin tests for the diagnosis of *M. tuberculosis* infection.

Individuals were injected intradermally with 100 μ l of either PBS or PBS plus Tween 20TM containing either 0.1 μ g of protein (for TbH-9 and TbRa35) or 1.0 μ g of protein (for TbRa38-1). Induration was measured between 5-7 days after injection, with a response of 5 mm or greater being considered positive. Of the 20 individuals tested, 2 were PPD negative and 18 were PPD positive. Of the PPD positive individuals, 3 had active tuberculosis, 3 had been previously infected with tuberculosis and 9 were healthy. In a second study, 13 PPD positive individuals were tested with 0.1 μ g TbRa11 in either PBS or PBS plus Tween 20TM as described above. The results of both studies are shown in Table 10.

TABLE 10
RESULTS OF DTH TESTING WITH REPRESENTATIVE ANTIGENS

	TbH-9 Pos/Total	Tb38-1 Pos/Total	TbRa35 Pos/Total	Cumulative Pos/Total	TbRa11 Pos/Total
PPD negative	0/2	0/2	0/2	0/2	
PPD positive					
healthy	5/9	4/9	4/9	6/9	1/4
prior TB	3/5	2/5	2/5	4/5	3/5
active	3/4	3/4	0/4	4/4	1/4
TOTAL	11/18	9/18	6/18	14/18	5/13

EXAMPLE 9

SYNTHESIS OF SYNTHETIC POLYPEPTIDES

hexafluorophosphate) activation. A Gly-Cys-Gly sequence may be attached to the amino terminus of the peptide to provide a method of conjugation or labeling of the peptide. Cleavage of the peptides from the solid support may be carried out using the following cleavage mixture: trifluoroacetic acid:ethanedithiol:thioanisole:water:phenol (40:1:2:2:3). After cleaving for 2 hours, the peptides may be precipitated in cold methyl-t-butyl-ether. The peptide pellets may then be dissolved in water containing 0.1% trifluoroacetic acid (TFA) and lyophilized prior to purification by C18 reverse phase HPLC. A gradient of 0%-60% acetonitrile (containing 0.1% TFA) in water (containing 0.1% TFA) may be used to elute the peptides. Following lyophilization of the pure fractions, the peptides may be characterized using electrospray mass spectrometry and by amino acid analysis.

EXAMPLE 10

PREPARATION AND CHARACTERIZATION OF *M. TUBERCULOSIS* FUSION PROTEINS

A fusion protein containing TbRa3, the 38 kD antigen and Tb38-1 was prepared as follows.

Each of the DNA constructs TbRa3, 38 kD and Tb38-1 were modified by PCR in order to facilitate their fusion and the subsequent expression of the fusion protein TbRa3-38 kD-Tb38-1. TbRa3, 38 kD and Tb38-1 DNA was used to perform PCR using the primers PDM-64 and PDM-65 (SEQ ID NO: 146 and 147), PDM-57 and PDM-58 (SEQ ID NO: 148 and 149), and PDM-69 and PDM-60 (SEQ ID NO: 150 and 151), respectively. In each case, the DNA amplification was performed using 10 μ l 10X Pfu buffer, 2 μ l 10 mM dNTPs, 2 μ l each of the PCR primers at 10 μ M concentration, 81.5 μ l water, 1.5 μ l Pfu DNA polymerase (Stratagene, La Jolla, CA) and 1 μ l DNA at either 70 ng μ l (for TbRa3) or 50 ng μ l (for 38 kD and Tb38-1). For TbRa3, denaturation at 94°C was performed for 2 min, followed by 40 cycles of 96°C for 15 sec and 72°C for 1 min, and lastly by 72°C for 4 min. For 38 kD, denaturation at 96°C was performed for 2 min, followed by 40 cycles of 96°C for 30 sec, 68°C for 15 sec and 72°C for 3 min, and finally by 72°C for 4 min. For Tb38-1 denaturation at

1.5 min. 30 cycles of 96°C for 15 sec, 64°C for 15 sec and 72°C for 1.5, and finally by 72°C for 4 min.

The TbRa3 PCR fragment was digested with NdeI and EcoRI and cloned directly into pT7⁺L2 IL 1 vector using NdeI and EcoRI sites. The 38 kD PCR fragment was digested with Sse8387I, treated with T4 DNA polymerase to make blunt ends and then digested with EcoRI for direct cloning into the pT7⁺L2Ra3-1 vector which was digested with StuI and EcoRI. The 38-1 PCR fragment was digested with Eco47III and EcoRI and directly subcloned into pT7⁺L2Ra3/38kD-17 digested with the same enzymes. The whole fusion was then transferred to pET28b - using NdeI and EcoRI sites. The fusion construct was confirmed by DNA sequencing.

The expression construct was transformed into BLR pLys S *E. coli* (Novagen, Madison, WI) and grown overnight in LB broth with kanamycin (30 µg/ml) and chloramphenicol (34 µg/ml). This culture (12 ml) was used to inoculate 500 ml 2XYT with the same antibiotics and the culture was induced with IPTG at an OD₅₆₀ of 0.44 to a final concentration of 1.2 mM. Four hours post-induction, the bacteria were harvested and sonicated in 20 mM Tris (8.0), 100 mM NaCl, 0.1% DOC, 20 µg/ml Leupeptin, 20 mM PMSF followed by centrifugation at 26,000 X g. The resulting pellet was resuspended in 8 M urea, 20 mM Tris (8.0), 100 mM NaCl and bound to Pro-bond nickel resin (Invitrogen, Carlsbad, CA). The column was washed several times with the above buffer then eluted with an imidazole gradient (50 mM, 100 mM, 500 mM imidazole was added to 8 M urea, 20 mM Tris (8.0), 100 mM NaCl). The eluates containing the protein of interest were then dialyzed against 10 mM Tris (8.0).

The DNA and amino acid sequences for the resulting fusion protein hereinafter referred to as TbRa3-38 kD-Tb38-1 are provided in SEQ ID NO: 152 and 153, respectively.

A fusion protein containing the two antigens TbH9 and Tb38-1 (hereinafter referred to as TbH9-Tb38-1) without a hinge sequence, was prepared using a similar procedure to that described above. The DNA sequence for the TbH9-Tb38-1 fusion protein is provided in SEQ ID NO: 154.

The ability of the fusion protein TbH9-Tb38-1 to induce T cell proliferation and IFN- γ production in PBMC preparations was examined using the protocol described above in Example 1. PBMC from three donors were employed: one who had been previously shown to respond to TbH9 but not Tb38-1 (donor 131); one who had been shown to respond to Tb38-1 but not TbH9 (donor 184); and one who had been shown to respond to both antigens (donor 201). The results of these studies (Figs. 5-7, respectively) demonstrate the functional activity of both the antigens in the fusion protein.

A fusion protein containing TbRa3, the antigen 38kD, Tb38-1 and DPEP was prepared as follows.

Each of the DNA constructs TbRa3, 38 kD and Tb38-1 were modified by PCR and cloned into vectors essentially as described above, with the primers PDM-69 (SEQ ID NO:150 and PDM-83 (SEQ ID NO: 205) being used for amplification of the Tb38-1A fragment. Tb38-1A differs from Tb38-1 by a *Dra*I site at the 3' end of the coding region that keeps the final amino acid intact while creating a blunt restriction site that is in frame. The TbRa3 38kD Tb38-1A fusion was then transferred to pET28b using *Nde*I and *Eco*R1 sites.

DPEP DNA was used to perform PCR using the primers PDM-84 and PDM-85 (SEQ ID NO: 206 and 207, respectively) and 1 μ l DNA at 50 ng/ μ l. Denaturation at 94 °C was performed for 2 min, followed by 10 cycles of 96 °C for 15 sec, 68 °C for 15 sec and 72 °C for 1.5 min, 30 cycles of 96 °C for 15 sec, 64 °C for 15 sec and 72 °C for 1.5 min; and finally by 72 °C for 4 min. The DPEP PCR fragment was digested with *Eco*R1 and *Eco*721 and clones directly into the pET28Ra3/38kD/38-1A construct which was digested with *Dra*I and *Eco*R1. The fusion construct was confirmed to be correct by DNA sequencing. Recombinant protein was prepared as described above. The DNA and amino acid sequences for the resulting fusion protein hereinafter referred to as TbF-21 are provided in SEQ ID NO: 208 and 209, respectively.

The reactivity of the fusion protein TbF-21 with sera from *M. tuberculosis*-infected patients was examined by ELISA using the protocol described above. The results of these studies (Table 11) demonstrate that all four antigens function independently in the fusion protein.

TABLE 11
REACTIVITY OF TbF-2 FUSION RECOMBINANT WITH TB AND NORMAL SERA

Serum ID	Status	TbF OD450	Status	TbF-2 OD450	Status	ELISA Reactivity			
						38 kD	TbRa5	Tb38-1	DPEP
B931-40	TB	0.57	-	0.321	-	-	-	-	-
B931-41	TB	0.601	-	0.396	-	-	-	-	-
B931-109	TB	0.494	-	0.404	-	-	-	-	-
B931-132	TB	1.502	-	1.292	-	-	-	-	-
5004	TB	1.806	-	1.666	-	-	-	-	-
15004	TB	2.862	-	2.468	-	-	-	-	-
39004	TB	2.443	-	1.722	-	-	-	-	-
68004	TB	2.871	-	2.575	-	-	-	-	-
09004	TB	0.691	-	0.971	-	-	-	-	-
107004	TB	0.875	-	0.732	-	-	-	-	-
92004	TB	1.632	-	1.394	-	-	-	-	-
97004	TB	1.491	-	1.979	-	-	-	-	-
118004	TB	3.182	-	3.043	-	-	-	-	-
173004	TB	3.644	-	3.578	-	-	-	-	-
175004	TB	3.332	-	2.916	-	-	-	-	-
274004	TB	3.696	-	3.716	-	-	-	-	-
276004	TB	3.243	-	2.56	-	-	-	-	-
282004	TB	1.249	-	1.234	-	-	-	-	-
289004	TB	1.373	-	1.17	-	-	-	-	-
308004	TB	3.708	-	3.355	-	-	-	-	-
314004	TB	3.663	-	3.309	-	-	-	-	-
317004	TB	1.763	-	0.92	-	-	-	-	-
312004	TB	1.709	-	1.453	-	-	-	-	-
380004	TB	0.238	-	0.461	-	-	-	-	-
451004	TB	0.18	-	0.2	-	-	-	-	-
478004	TB	0.188	-	0.469	-	-	-	-	-
410004	TB	0.384	-	2.392	-	-	-	-	-
411004	TB	0.306	-	0.874	-	-	-	-	-
421004	TB	0.357	-	4.56	-	-	-	-	-
528004	TB	0.047	-	0.196	-	-	-	-	-
A6-87	Normal	0.094	-	0.063	-	-	-	-	-
A6-88	Normal	0.214	-	1.19	-	-	-	-	-
A6-89	Normal	0.248	-	0.125	-	-	-	-	-
A6-90	Normal	0.179	-	0.296	-	-	-	-	-
A6-91	Normal	0.133	-	0.151	-	-	-	-	-
A6-92	Normal	0.064	-	0.097	-	-	-	-	-
A6-93	Normal	0.072	-	0.098	-	-	-	-	-
A6-94	Normal	0.072	-	0.064	-	-	-	-	-
A6-95	Normal	0.125	-	0.159	-	-	-	-	-
A6-96	Normal	0.121	-	0.12	-	-	-	-	-

A fusion protein containing TbRa3, the antigen 38kD, Tb38-1 and TbH4 was prepared as follows.

Genomic *M. tuberculosis* DNA was used to PCR full-length TbH4 (FL TbH4) with the primers PDM-157 and PDM-160 (SEQ ID NO: 348 and 349, respectively) and 2 μ l DNA at 100 ng/ μ l. Denaturation at 96 °C was performed for 2 min, followed by 40 cycles of 96 °C for 30 sec, 61 °C for 20 sec and 72 °C for 5 min; and finally by annealing at 72 °C for 10 min. The FL TbH4 PCR fragment was digested with EcoRI and Sca I (New England Biolabs.) and cloned directly into the pET28Ra3/38kD/38-1A construct described above which was digested with DraI and EcoRI. The fusion construct was confirmed to be correct by DNA sequencing. Recombinant protein was prepared as described above. The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbF-6) are provided in SEQ ID NO: 350 and 351, respectively.

A fusion protein containing the antigen 38kD and DPEP separated by a linker was prepared as follows.

38 kD DNA was used to perform PCR using the primers PDM-176 and PDM-175 (SEQ ID NO: 352 and 353, respectively), and 1 μ l PET28Ra3/38kD/38-1/Ra2A-12 DNA at 110 ng/ μ l. Denaturation at 96 °C was performed for 2 min, followed by 40 cycles of 96 °C for 30 sec, 71 °C for 15 sec and 72 °C for 5 min and 40 sec; and finally by annealing at 72 °C for 4 min. The two sets of primers PDM-171, PDM-172, and PDM-173, PDM-174 were annealed by heating to 95 °C for 2 min and then ramping down to 25 °C slowly at 0.1 °C/sec. DPEP DNA was used to perform PCR as described above. The 38 kD fragment was digested with Eco RI (New England Biolabs) and cloned into a modified pETAL2 vector which was cut with Eco 72 I (Promega) and Eco RI. The modified pETAL2 construct was designed to have a MGHHHHHH amino acid coding region in frame just 5' of the Eco 72 I site. The construct was digested with Kpn 2I (Gibco, BRL) and Pst I (New England Biolabs) and the annealed sets of phosphorylated primers (PDM-171, PDM-172 and PDM-173, PDM-174) were cloned in. The DPEP PCR fragment was digested with Eco RI and Eco 72 I and cloned into this second construct which was digested with Eco 47 III (New England Biolabs) and Eco RI. The resulting construct was then transformed into *E. coli* BL21 (DE3) and

modified pET28 vector. The fusion construct was confirmed to be correct by DNA sequencing.

Recombinant protein was prepared essentially as described above. The DNA and amino acid sequences for the resulting fusion protein (hereinafter referred to as TbF-8) are provided in SEQ ID NO: 354 and 355, respectively.

One of skill in the art will appreciate that the order of the individual antigens within the fusion protein may be changed and that comparable activity would be expected provided each of the epitopes is still functionally available. In addition, truncated forms of the proteins containing active epitopes may be used in the construction of fusion proteins.

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for the purpose of illustration, various modifications may be made without deviating from the spirit and scope of the invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANTS: Reed, Steven G.
Skelky, Yasir A.W.
Dillon, Davin C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald

(ii) TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS

(iii) NUMBER OF SEQUENCES: 355

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: SEED and BERRY LLP
(B) STREET: 6300 Columbia Center, 701 Fifth Avenue
(C) CITY: Seattle
(D) STATE: Washington
(E) COUNTRY: USA
(F) ZIP: 98104 0092

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.3, Version #1.30

vi. CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE: 05-MAY-1998
(C) CLASSIFICATION:

viii. ATTORNEY/AGENT INFORMATION:

A. NAME: Mark David E.
B. REGISTRATION NUMBER: 11,100
C. REFERENCE DOCUMENT NUMBER: 0111141109

ix. TELECOMMUNICATION INFORMATION

(A) TELEPHONE: 206 522-4900
(B) TELEFAX: 206 522-6931

x. INFORMATION FOR SEQ ID NO. 1

a. SEQUENCE CHARACTERISTICS:

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO 1:

CGAGGCACCG	GTAGTTTGAA	CCAAACGCAC	AATCGACGGG	CRAACGAACG	GAAGAACACA	60
ACCATGAAGA	TGGTGAAATC	GATCGCCGCA	GGTCTGACCG	CCGCGGCTGC	AATCGGCGCC	120
GCTGCGGCCG	GTGTGACTTC	GATCATGGCT	GGCGGCCCGG	TCGTATACCA	GATGCAGCCG	180
GTCTCTTCG	GCGCGCCACT	GCCGTTCGAC	CCCGCATCCG	CCCCTGACGT	CCCGACCGCC	240
GCCCAGTTGA	CCAGCCTGCT	CAACAGCCTC	GCCGATCCCA	ACGTGTCGTT	TGCGAACAAG	300
GGCAGTCTGG	TCGAGGGCGG	CATCGGGGGC	ACCGAGGCGC	GCATCGCCGA	CCACAAGCTG	360
AAGAAGGCCG	CCGAGCACCG	GGATCTGCCG	CTGTCGTTCA	GCGTGACGAA	CATCCAGCCG	420
GCGGCCGCCG	GTTCCGGCCAC	CGCCGACGTT	TCCGTCTCGG	GTCCGAAGCT	CTCGTCGCCG	480
GTCAACGAGA	ACGTCAAGTT	CGTGAATCAA	GGCGGCTGGA	TGCTGTACAG	CGCATCGGCG	540
ATGGAGTTGC	TGCAGGCCGC	AGGGNAACTG	ATTGGCGGGC	CGGNTTCAGC	CCGCTGTTCA	600
GCTACGCCGC	CCGCCTGGTG	ACCGGTCCAT	GTCAACACT	CGCGCGTGTA	GCACCGTGCG	660
GTNTGCCGAG	GGNCCACGCG	ACCGCCCGGT	GCAAGCCGTC	CTCGAGATAG	GTGGTGNCTC	720
GNCACGAGNG	ANCACCCCTN	MNTCGNCINT	TCTCGNTGNT	GNATGA		766

(2) INFORMATION FOR SEQ ID NO:1:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 762 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ATGCATCACC	ATCACCATCA	CGATGAAGTC	ACGGTAGAGA	CGACCTCCGT	CTTCCGCGCA	60
GACTTCCTCA	GCGAGCTGGA	CGCTGCTGCG	CAAGCGGGTA	CGGAGAGCGC	GGTCTCCGGG	120
GTGGAAGGGC	TCCCGCCCGG	GTGCGCGTTG	GTGGTAGTCA	AACGAGGCCG	CAACGCCGGG	180
TCCCGGTTCC	TACTCGACCA	AGGCATCAGC	TGCGGTGGTC	GCGATCCCGA	CAGCGACATA	240
TTCTGCGACG	ACGTGACTGT	GAGCGGTGCG	CATGCTGAAT	TCCCGTTGGA	AAACAACGAA	300
TTCAATGTCC	TGGATGTCCG	GAGTGTGAAT	GCGACCTAGC	TCAACCGCGA	GCCCCGTGAT	360
TGGGCGGTGC	TGGCGAACGG	CGACGAGGTC	CAGATCGGCA	AJCTCCGCTT	GGTGTTCCTG	420
ACTGGACCCA	AGCAACGCCA	GGATGACGGG	AGTACGCGCG	GCCCGTGAGC	GCACCCGATA	480
GCCCCCGGCT	GCCCCGGGATG	TCGATCGGGG	CGGTCTCTCG	ACCTGCTACG	ACCGGATTTT	540
CGCTGATGTC	CACCATCTCC	AAGATTGCGT	TCTTGGGAGG	CTTGAGGGTC	NGGGTGACCC	600
CGCCGCGGGC	CTCATTGNGG	GSTNTCGGCH	GGTTTGACCG	CTTACCTACT	GCCTCCCGGN	660
TTGCAATATC	TTTCTTCTCT	GGCCTTAAAT	GGATCTTAA	CTTCCCGGCT	GAAGGGTNA	720
TTTGGGGGCT	TTTCTTGAAT	GGCCTTAAAT	GGATCTTAA	CTTCCCGGCT	GAAGGGTNA	766

(2) INFORMATION FOR SEQ ID NO:2:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 813 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CGGTATGAAC	ACGGCCCGCT	CGGATAACTT	CCAGCTGTCC	CAGGGTGGGC	AGGGATTCCG	60
CATTCCGATC	GGGCAGGCGA	TGGCGATCCG	GGGCCAGATC	CGATCGGGTG	GGGGGTCAAC	120
CACCGTTGAT	ATCGGGGCGTA	CCCGCTTCCG	CGGTTTGGGT	GTGTGTCGAC	ACAACGGCAA	180
CGGCGCACGA	GTCCAAACGG	TGTTCTGGGAG	CGCTCCGGCG	GCAAGTCTCG	GCATCTCCAC	240
CGGCGACCGT	ATCACCCCGG	TGACGGGCGG	TCCGATCAAC	TCGGCCACCG	CGATGGCGCG	300
TGGCGTTAAC	GGGCATCATC	CGGCTGAGGT	CATCTCGGTG	AACTGGCAAA	CGAAGTCCGG	360
TGCGACCGCT	ACAGTGAACG	TGACATCGGC	CGAGGGACCG	CGGGCCTGAT	TTCTGTCYGG	420
ATACGACCGG	CGCGCCGGCG	AATTGGA				447

TTTGGGTTT	TGTGGGGAG	TATGTGGG	ATTAATTTT	TGGGAGGGG	GGAAATTAAT	120
TTTGTATTT	TAGTTGGG	TTTGTGGAA	TTTGTGGG	TTAGGTATG	TTGGATGGG	120
AGGGGGGGGA	GGGGGAGGG	GGGAATGGG	GGATGAGGA	TGGGGGCAAT	TTGGGGGGGG	180
TTGGGGAGG	TTAGGGGGGG	AATGGGGGA	GTGGGAGGT	GGGAGGTCT	GGGGAGGTT	240
TTGGGATGAA	TGTGATTTG	TTTGGGGG	GGATTTGGA	ATGGAGGTAG	TGAGGGGAAA	300
TTGATGATG	AAAAAGGGG	GGGAGGTGG	TTGTTGTTT	GGTGTAGGT	GGGTTGGTGG	360
TTTGGGGTT	ATGAGGATG	TTTGGGGG	AAGGTGATG	GGAGGAACAG	GGTGTGGGG	420
TTTATTTGAA	TTGTTGTAAT	TTTGGGGG	TTGGGATG	TTTATGAGG	TTTGGTGGG	480
TTTATTTG	TTTATTTG	TTTGGGGG	TTTATTTG	TTTATTTG	TTTATTTG	540

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 633 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TTGCANGTCG AACCACCTCA CTAAAGGGAA CAAAAGCTNG AGCTCCACCG CGGTGGCGGC	60
CGCTCTAGAA CTAGTGKATM YYYCKGGCTG CAGSAATYCG GYACGAGCAT TAGGACAGTC	120
TAACGGTCTCT GTTACGGTGA TCGAATGACC GACGACATCC TGCTGATCGA CACCGACGAA	180
CGGGTGCGAA CCTCACCCT CAACCGGCCG CAGTCCCGYA ACGCGCTCTC GCGCGCGCTA	240
CGGGATCGGT TTTTCGCGGY GTTGGYCGAC GCCGAGGYCG ACGACGACAT CGACGTCTGC	300
ATCCTCACCG GYGCGGATCC GGTGTTCTGC GCCGGACTGG ACCTCAAGGT AGCTGGCCCG	360
GCAGACCGCG CTGCGGACA TCTCACCCTG GTGGCGGGCC ATGACCAAGC CGGTGATCGG	420
CGCGATCAAC GCGCGCGCGG TCACCGCGCG GCTCGAAGTG GCGGTGTACT GCGACATCCT	480
GATCGCTTCC GAGGACGCGG GCTTCGNCGA CACCGACGCG TGGGTGGCGG TGCTGCGGAC	540
TTGGGAGCTC AGTGTGTGCT TCGCGGAAAA GGTGGGATC GGNATGGCGG GGTGGATGAG	600
CGTGACCGGG GACTACCTGT CGTGACCGGA CGC	633

(x1) INFORMATION FOR SEQ ID NO 7:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1362 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGACGACGAC GCGGCGGAG AGCGGGCGCG AACGGCGATC GACCGCGGCC TGGCGAGAGT	60
TGCGACCACT GAGGAGGGAG TCGAATCATG AAATTGTGCA ACCATATTGA GCGCGTGGCG	120
TTGGCGGAG GCGGCGGCGG GGTGGCGGAG GTTATGCGCG AGCGCGCGCG CGAGTTCGGG	180
TGGGTGGCGG AGTGGCTGCG CATGCTGTGC TTGACGAGG GATGCTCTAC CGCGGGCTGG	240
CGACGCTTGC CGGAGAGACT GCTGGTGGGC GAGGTGGCGG GTGGCGGCAA GGAAGCGCTC	300
GCGCGCGCGG TGGCGGCGAG GCTGGCTGCG TGGTGGTGGG TGACCGCACA CACCACCATG	360
CTGTACGCGG CAGGCGCAAA GACACCGGCG GCGCGGATCT TGGCGGGCAC AGCACCTGCC	420
GCGGGTGACC CGAAGCGGCT GATGTGGCG TGGCGGGCAG GAACCGGGAC ACCGGCGGGA	480
TGCGCGGCAC GTTGGCGGCG GATGTGGCG TGGGAATAGG TGGCGAGCGG GGTGCAATTG	540
GACTTCATCG GACTGCTGGT GCTGGTGGTG TTGAGGAAA GTTGGCTGCG GGGGGCGCGG	600
GCGCGGCAAG AGCTCATGCG GCGCGGCGGT TGAGTGGTGT TGGCGGCAA GGTGGCGCGG	660
GAGCATGGCG GGGCGGCTG GACCGGCGCG TTGAGGCTG GAAGGCTGCG CGACGATCTG	720
GATGGGCAA GAGGCTGGA GCGCATAGA ACCGCTTCT GCGGCTGAG CACACCTG	780
GACACCGCGG GCGAGCTGCG GCGACGAGT GCTAGGTTGG TCGGGCGGCT GGTGGGTTG	840
TGCGAGCGCG AGGCAATGCG GATGAGAGT GCTGGAGCA ACGAGCACAC CCGCGAGCTG	900
GCGGGGAGCG TGCGAGCGCG CACCGCTGTT GCGGTGCTGA GCGGCTGGG CCGCGATCAG	960
GTGAGCGAGC AGGAGCTGCG GCGGGCGGCA TGGGTGCTG ACAGGAGTGC GCGGCTGGTT	1020
TGCGCGCTGG GCTGGGCGCG GTTACCGCG GCGCGGCGCA TGCGGAGCTG GATGCGCGCG	1080
GCGCGCGAGG GCGAGGTGCG GCGGCAAAA CCGACTGGGT GAGTGTGGCG GCGCTGTGCG	1140
TAGGCTGTCA TGTGTGGCG GAGGATCTG TCGGGCTGA AGTGAATG GACACAGGT	1200
TAGGCTGTCA TGTGTGGCG TCGGGCTGA TCGGGCTGA TCGGGCTGA TCGGGCTGA	1260

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1458 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```

GCGACGACCC CGATATGCCG GGCACCGTAG CGAAAGCCGT CGCCGACGCA CTCGGGCGCG 60
GTATCGCTCC CGTTGAGGAC ATTCAGGACT GCGTGGAGGC CCGGCTGGGG GAAGCCCGTC 120
TGGATGACGT GGGCCGTGTT TACATCATCT ACCGGCAGCG GCGCGCCGAG CTGCGGACGG 180
CTAAGGCCCTT GCTCGGCGTG CGGGACGAGT TAAAGCTGAG CTTGGCGGCC GTGACGGTAC 240
TGCSCGAGCG CTATCTGCTG CACGACGAGC AGGGCCGGCC GGGCGAGTCG ACCGGCGAGC 300
TGATGGACCG ATCGGCGCGC TGTGTGCGCG GGGCGAGGA CCAGTATGAG CCGGGCTCCT 360
CGAGGCGGTG GGGCGAGCGG TTCGCCACGC TATTACSCAA CCTGGAATTC CTGCCGAATT 420
CGCCACCGTT GATGAACCTT GGCACCGAGC TGGGACTGCT GCGCGGCTGT TTTGTTCTGC 480
CGATGAGGA TTGCTGCAA TCGATCTTTG GACGCTGGG ACAGGCGCGC GAGCTGCGAC 540
GGGCTGGAGG CGGCACCGGA TATGCGTTCA GCGACCTGGC ACCCGCGCGG GATCGGGTGG 600
GCTCCACGGG CGGCACCGCC AGCGGACCGG TGTGTTTTCT ACCGCTGTAT GACAGTGGCG 660
CGGGTGTGGT CTCCATGGGC GGTGCGCGGC GTGGCGCCTG TATGGCTGTG CTTGATGTGT 720
CGCACCCCGA TATCTGTGAT TTGTCACCG CCAAGGCCGA ATCCCCCAGC GAGCTCCCGC 780
ATTCAACCT ATCGGTGGT GTGACCGAGC CTTCTCTGCG GCGCGTCGAA CGCAACGGCG 840
TACACCGGCT GGTCAATCGG CGAACCGSCA AGATCGTCCG GCGGATGCCC GCGCGCGAGC 900
TGTTCGACCG CATCTGCAA GCGCGSCAGC CCGGTGGCGA TCGCGGGCTG GTGTTTCTCG 960
ACACGATCAA TAGGGCAAAC CCGGTGCGCG GGAGAGGCGG CATCGAGCGC ACCAAGCCCT 1020
CGCGGGAGGT CCGACTGCTG CTTACGACT GTCGACTGGG ACCGGCTCGA GGAGGTGCGC GGTGTGGCGG 1080
CGATGCTCGC CGACGCTCGG GTGACTGGG GCGGCTACCG CTTCCCTCGAA CTGGGTGAGG 1140
TGGGTTGCTT TGATGACGTC ATCGATGTCA GCGGCTACCG CTTCCCTCGAA CTGGGTGAGG 1200
CGGCGCGCGC CACCCGCAAG ATCGGGTGG GACTCATGGG TTTGGTGGAA CTGCTTGGCG 1260
CAGTGGGTAT TCGGTACGAC AGTGAAGAAG CCGTGGCGTT AGCGACTCGG CTCATGCGTC 1320
GCATACAGCA GCGCGCGCAC ACGGCTCGG CGAGGCTGGC CGAAGAGCGG GCGCGATTCC 1380
CGGCTTCGCA CGATAGCGCG TTGCGCTGCT TCGCGCGGAG TGTCAAGCA CAGGTCACT 1440
CGGTGCTGCT GACGGGCA

```

2. INFORMATION FOR SEQ ID NO:4

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 962 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

1. SEQUENCE DESCRIPTION: SEQ ID NO:4

```

ACGCTTAAT CTTCTGAT CTGGAACCT TTGCGGCT ACCTACGAG ATCTACTGGC 60
CGCCGAGGG CTTGGCTGTC GGCATCGCG TGTCTGAGT CGGATCGCG GTGGCATCG 120
TCATCGGCTT CTTGACAGG AGCCCGGTC CCAAACTGCT CAGCGCGAG AAGTGGGCT 180
CGCCGAGAG CCGTGGGGT TCGCGGCA TTAAGTAC CAGCGGCTT GGTCAACCG 240
AACTAACCT CCGTGGGGT TCGCGGCA TTAAGTAC CAGCGGCTT GGTCAACCG 300
CGGCTTCGA CCGTGGGGT TCGCGGCA TTAAGTAC CAGCGGCTT GGTCAACCG 360

```

GGACCGGGAT GGGATCGGCG CGCGCTGCGG CATTCGCCCG GCGGGCGATC GGGCGGGGCA	660
CCTACAATCT CGTGGTACAA CTGGGCAATC TCGGCTCGCT GCGGTTCCG TTCATCCTGA	720
ATCAGCCGCG GCGCGCGCGG GGGCGGTAC CGGCTCGGG TCCAGCGCAG GCGCCTCCGC	780
CGGAGTCTCC CGCGCAAGGC GGATAATTAT TGATCGCTGA TGGTCGATTC CGCCAGCTGT	840
GACAACCCCT CGCCTCGTGC CG	862

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 622 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TTCATCAGCA CCGGCAAGGC GTCACATGCC TCGCTGGTGC TCGAGGTGAC CAATGACAAA	60
TACACCCCGG GCGCCAAGAT CGTCAAGTA GTGGCCCGTG GTGCTGCCCG GAACCGTGA	120
GTGGCGAAGG CGCTCGTTGT CACCAAGGTC GACGACCGCG CGATCAACAG CGCGGACCGG	180
TTGGTTGCGG CGCTCGCGTC CAAAGCGCGG CGCGCCACGG TGGCGCTAAC CTTTCAGGAT	240
CGCTCGGGCG GTAGCCCGCAC AGTGCAAGTC ACCCTCGGCA AGGCGGAGCA GTGATGAAGG	300
TGGCGCGCA GTGTTCAAAG CTCGGATATA CGGTGGCACC CATGGAACAG CGTGGCGAGT	360
TGGTGGTTGG CCGGGCACTT GTCGTCTGCG TTGACGATCG CACGGCGCAC GCGGATGAAG	420
ACCACAGCGG GCGGCTTGTG ACCGAGCTGC TCACCGAGGC CGGGTTTGTG GTCGACGGCG	480
TGGTGGCGGT GTCGGCGCAC GAGGTGAGAG TCCGAAATGC GCTGAACACA GCGGTGATCG	540
GCGGGGTGGA CGTGGTGGTG TCGGTGCGCG GGACCGNGT GACGNTCTGC GATGTCACCC	600
CGGAAGCCAC CGNGACATT CT	622

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1200 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ATGCGCGCGG TAAGCGTGTG TCGCGCGCGG ACCTCGGTGT TCACAGCATC CGGCGGTGCG	60
ATGACAGCTT CGTCTCAGG TCGCGCGCGA ACCTCGGTGT CGTTCAGTTC CGGCGCGAAG	120
AAAGAGCTCC AGTTCAGCGG TCGACCGCGA AAAGAAATG CCATCGAGCA GTTCGTGTAT	180
ATGACCTGC GATCGTCTCG TCGTACAGG TCGGACTACA AGCGCAACGG GTCCGCTGCG	240
ATGTTGACCC AGTTTCTCAA AAAGAAAGC GATTCGCGG GTTCGATGT CCGGTTGAAT	300
ATGTCGACCG GTCAAGCTGA TCGGTCGCGG TCGGCTGCG GTTCGCGCGG ATGGGACGTC	360
ATGAGGCTGT TCGGCGCGAT TCGATCAGT TACAAATATCA AGGGCGTGAAG CACGCTGAAT	420
ATGACGAGAC CGCTACCGT TAAGATTTC AAAGGAGCA TCACCGTGTG GAATGATCA	480
TATATCCAG CGCTCACTC CGGACCGAG GTCGCGCGAA CACCGATTAG CGTTATCTTC	540
TCAGGAGACA AGTCCGCTAC GTCGACAAAC TTCAGAAAT ACCTCGACCG TGTATCCAAC	600
ATGGGCTGCG CCAAAGCGCG TAGGAAAGC TTCAGCGGGG GGTTCGCGCT CGGCGCGAGG	660
ATGAAAGAGG GAACTCTGGT TATATCCAG ATGACGAGG GTTCGATCA CTACAAGTAA	720
CGCTCTCTTC GGTTCGCTAA TATGTAAT ATGCGGAGG TATATCTTC GTCGCTGCGG	780

GACCAATACG GGTCCATTCC GTTGCCCAAA TCGTTCCAAG CAAAATTGGC GGCCGCGGGT	1080
AATGCTATTT GTTGACCTAG TGAAGGGAAT TCGACGGTGA GCGATGCCGT TCCGCAGGTA	1140
GGGTGCGCAAT TTGGGGCCGT TCAGCTATTG CGGCTGCTGG GCCGAGGCGG GATGGGCGAG	1200

(2) INFORMATION FOR SEQ ID NO:12:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1155 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GCAAGCAGCT GCAGGTGCTG CTSTTCGACG AACTGGGCAT GCCGAAGACC AAACGCACCA	60
AGACCGGCTA CACCACGGAT GCCGACCGCG TGCAGTCGTT GTTCGACAAG ACCGGGCATC	120
CGTTTCTGCA ACATCTGCTG GCCGACCGCG ACCTCAGCGG GCTCAAGGTC ACCGTGACCG	180
GGTTGCTGCA AGCGGTGCGC GCGGACCGCG GCATCGACAC CACGTTCAAC CAGACGATCG	240
CGCGCAACCG CGCGCTGCTG TCGACCGAAC CCAACCTGCA GAACATCCCG ATCCGCGACCG	300
ACGCGGGCGG CGCGATCCCG GACGCTTCCG TGGTCGGGGA CGGTTACCGC GAGTTGATGA	360
CGGCGGACTA CAGCCAGATC GAGATGCGGA TCATGGGGCA CCTGTCCGGG GACGAGGGCG	420
TCATCGAGGG GTTCAACACC GGGGAGGACC TGTATTGCTT CGTCCGCTCC CGGGTGTTCG	480
GTGTGCCCCA CGACGAGGTC ACCGGCGAGT TCGCGCGCGG GGTCAAGCGG ATGTCTTACG	540
GGCTGGTTTA CGGTTTCAGC GCCTACGGCG TGTGCGAGCA GTTGAAAATC TCCACCGAGG	600
AAGCCCAACG SCAGATCGAC GCGTATTTCG CCGGATTCGG CGGGGTGCGC GACTACCTGC	660
CGGCGCTAGT CGAGCGGGCG CGCAAGGACG GCTACACCTC GACGGTGCTG GGCCGTCGCG	720
GCTACCTGCG CGAGCTGGAC AGCAGCAACC GTCAAGTCCG GGAGGCGCGC GAGCGGGCGG	780
CGCTGAACCG CGCGATCCAG GCGAGCGCGG CCGACATCAT CAAGGTGGCG ATGATCCAGG	840
TCGACCAAGG GGTCAACGAG GCACAGCTGG CGTCGCCCAT GCTGCTGCAG GTCCACGACG	900
AGCTGCTGTT CGAAATCCCG CCGGTTGAAC GCGAGCGCGT CGAGGCGCTG GTGCGCGACA	960
AGATGGGCGG CGCTTACCGG CTCGACGCTC CGCTGAGGT CTCGGTGGGC TACGGCGCGCA	1020
GCTGGGACCG CGCGGCGGAC TGAGTGGCGA GCTGCTATC CGGGCGGGAA TTCGGCGATT	1080
CTTGGCGCGT GAGTTCAAGC TCGCGCGCAAT CGGACCGAG TTTCTGAGC GTGTACCGCT	1140
CGAGTAGCGT CTCA	1155

(2) INFORMATION FOR SEQ ID NO:13:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1171 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GAGTTCGCTG TGGTGTGTTA AGGTTTTCAG CGGTGGGCAT CGGACCGCGG GTTCTCGGGT	60
TGGTGTGCTG GGTGCGGAT GTTCAACAG GTGTTGCTTA AGCAGCGCGG ATTGCTGCGC	120
ATCGAAGACA CGGACCGAGG TGGGACCGCG CCTGGAACGT CGATTACCT GCTGCTGCTG	180
CGGCTGCGGA TCGGATTCG GAGCTTTCG GGTGCGACGA GTGCGCTCG GAGTACGTA	240
ATCGAAGAT CTGGGGCTTC GCGAAAGCT ATTGAGTGA AATCTGAGTC CAGGCTGCGA	300
AATGATGCTT CGAGTTACTG TCGAAGACA GAGGCTGCTG GAGGCTGCTG GAGGCTGCTG	360

GACCATGACG	CCCCCTCCTG	GGATGGTTGG	CCAAACGCCCT	CGTGCAAGGCA	TGTTGGCCAT	660
CGGCGCGGTG	ACGATAGCGG	TGGTGTCCGG	CGGCATCGGC	GGCGCGGCCG	CATCCCTGGT	720
CGGGTTCAAC	CGGGCACCCG	CCGGCCCCAG	CGGCGGCCCA	GTGGCTGCCA	GCGCGGCGCC	780
AAGCATCCCC	GCAGCAAACA	TGCCGCCGGG	GTCCGGTCGA	CAGGTGGCGG	CCAAGGTGGT	840
GCCCAGTGTC	GTCTGTGG	AAACCGATCT	GGGCGGCCAG	TCCGAGGAGG	GCTCCGGCAT	900
CATTCTGTCT	GCCGAGGGGC	TGATCTTGAC	CAACAACCAC	GTGATCGCGG	CGGCCGCCAA	960
GCCTCCCTTG	GGCAGTCCGC	CGCCGAAAAC	GACGGTAACC	TTCTCTGACG	GGCGGACCGC	1020
ACCCCTTACG	GTGGTGGGGG	CTGACCCAC	CAGTGATATC	GCCGTCTGCC	GTGTTGAGGG	1080
CGTCTCCGGG	CTCACCCCGA	TCTCCCTGGG	TTCTCTCTCG	GACCTGAGGG	TCCGTGAGCC	1140
GCTGTGGCGG	ATCGGGTCCG	CGCTCGGTTT	GGAGGGCACC	GTGACCACGG	GGATCGTCAG	1200
CSCTCTCAAC	CGTCCAGTGT	CGACGACCGG	CGAGGCCGGC	AACCAGAACA	CCGTGCTGGA	1260
CGCCATTTCAG	ACCGACGCCG	CGATCAACCC	CGGTAACCTC	GGGGCGCGCG	TGGTGAACAT	1320
GAACGCTCAA	CTCGTCCGAG	TCAACTCGGC	CATTGCCACG	CTGGGCGCGG	ACTCAGCCGA	1380
TGCGCAGAGC	GGCTCGATCG	GTCTCGGTTT	TGCGATTCCA	GTGACCCAGG	CCAAGCGCAT	1440
CGCCGACGAG	TTGATCAGCA	CGGCAAGGC	GTCCATATGC	TCCCTGGGTG	TGCAGGTGAC	1500
CAATGACAAA	GACACCCCGG	CGCCCAAGAT	CGTCCAAGTA	GTGGCCCGGT	GTGCTGCCCG	1560
GAACCCCTGA	GTGCCGAAGG	CGCTCGTTGT	CACCAAGGTC	GACGACCCGC	CGATCAACAG	1620
CGCGGACCGG	TTGGTTGCCG	CGCTGCCGTC	CAAAGCGCGG	GGGCGCACCG	TGGCGCTAAC	1680
CTTCAGGAT	CCCTCGGGCG	GTAGCCGCAC	AGTGCAAGTC	ACCCCTCGGC	AGGCGGAGCA	1740
GTGATGAAGG	TCCCGCGCGA	GTGTTCAAAG	C			1771

(2) INFORMATION FOR SEQ ID NO:14:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1058 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CTCCACCGCG	GTGGCGCGCG	CTCTAGAACT	AGTGGATCCG	CGGGGCTGCA	GGAAATCGGC	90
ACGAGGATCC	GACGTCCGAG	CTTGTGGAAC	TCGCGCGCGC	CGAAGTATCG	GTCCATGCTT	120
AGCCCGCGGA	CGGCGAGCGC	CGGAATGGCG	CGAGTGAGGA	CGCGGGCAAT	TTGGCGGGGC	180
CGCGCGACCG	CGAGCGCGCG	AATGGCGCGA	GTGAGGAGGC	CGCCAGTCAT	GCCGAGCGTG	240
ATCCAAATCAA	CCTGCATTCC	GCCTGCGGGC	GCATTTGACA	ATCCAGGTAG	TGAGCGCAAA	300
TGAATGATGG	AAAACCGGCG	GTGACGTCCG	GTGTTCTGGT	GGTGCTAGGT	GCCTGCCTGG	360
CGTTGTGGCT	ATCAGGATGT	TCTTCGCGCA	AACCTGATGC	CGAGGAACAG	GGTGTTCGCG	420
TGACCCCGAC	GCCCTCCGAC	CGCGCGCTCC	TCTCCAGAT	CAGGCAGTCC	CTTGATGCGA	480
TAAGGGGTT	GACGAGCTTG	CTCCTAGCGG	TCTCAACAC	CGGGAAGTCC	GACAGCTTGC	540
TGGGTATTAC	CAGTCCCGAT	GTCCACCTCC	GGGCAATCC	CTCCTCGCGA	AAGGGCGTAT	600
GCACCTACAA	CGACGAGCGG	GTGTCCCGCT	TTGCGGTACA	AGGCGACAAC	ATCTCGGTGA	660
AACTGTTCCA	CGACTGGAGT	AACTCGCGCT	CGATTTCTGA	ACTGTCAACT	TCACGCGTGC	720
TGATCCCTCC	CGCTGGGCTG	AAGCAGCTGC	TCTCGGGTGT	CACGAACCTC	CAAGCGCAAG	780
GTACCGAAGT	GATAGACCGA	ATTTCGACCA	CGAAAATCAC	CGGAGCCATC	CGCGCGAGCT	840
GTGTCAAGAT	CGTTGATCTT	GGGCGCAAGA	GTGCAAGGCT	GGTGAACCTG	TGGATTGCCC	900
AGGACCGGTC	GCACGAGCTC	GTCCGAGCGA	CTCTGACCTT	CGATCCCGGG	TGATTCAGCG	960
TGACCGCAGT	GAAATGGAAC	GAACCCCTCA	ACCTGACTA	CGTGAAGTT	CGCTCGAGCG	1020
GTGNTGCGAA	ACGCGCTTCT	GAAGGTCTT	AAAGGNA			1058

(2) INFORMATION FOR SEQ ID NO:14:

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAATTCGGCA CGAGAGGTGA TCGACATCAT CGGGACCAGC CCCACATCCT GGGAAACAGGC	60
GGCGGGCGGAG GCGGTCCAGC GGGCGCGCGA TAGCGTCGAT GACATCCGCG TCGCTCGGGT	120
CATTGAGCAG GACATGGCCG TGGACAGCCG CGGCAAGATC ACCTACCGCA TCAAGCTCGA	180
AGTGTCGTTT AAGATGAGGC CGGCGCAACC GCGCTAGCAC GGGCCGGCGA GCAAGACGCA	240
AAATCGCAGC GTTTGCGGTT GATTGCTGCG ATTTTGTGTC TGCTCGCCGA GGCCTACCAG	300
GCGCGGCCCA GGTCCGCGTG CTGCCGTATC CAGGCGTGCA TCGCGATTCC GCGGCGCCAG	360
CCGGAGTTAA TGCTTCGCGT CGACCCGAAC TGGGCGATCC GCGGNGAGC TGATCGATGA	420
CCGTGGCCAG CCGTTCGATG CCGAGTTGC CCGAGGAAAC GTGCTGCCAG GCCGGTAGGA	480
AGCGTCCGTA GCGGCGCGTG CTGACCGGCT CTGCCTGCGC CCTCAGTGCG GCCAGCGAGC	540
GG	542

(x) INFORMATION FOR SEQ ID NO:16:

1. SEQUENCE CHARACTERISTICS:

(A) LENGTH: 913 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CGGTGCGGCG GCGGCGGCGG TTGCGGCGAT TGCGGCGGTC GCGGATCAGC TGCGCATGCG	60
CACCATCACC GCGTTGCGCG GCGGCGGCGG CGGTGCGGCG GGGGCGGCGG ATGCCACCGC	120
TTGACCGCTG GCGGCGGCGG CGGCATTGCG ATACAGCAGC CCGCGGGGGG CACCGTTACC	180
GCGGTGCGCA CCGTTCGCGG CGGTGCGGTT TCAGGCGGCG GAGGCGCGAAT GAACCGCGCG	240
CAAGCGCGCG GCGGCGGCGG TTGCGGCGTT TTGCGGCGCG CCGCGCGGCG CCGCCAATTG	300
CCGAACAGCG AMGCACGCTT GCGGCGGCGG CCGCGGCGGT TAACGCGCGT GCGGCGGCGG	360
GCGGCGGCGG CCGGCGGCGG GCGGCGGCGG GTTCGCGTGC CCGCGGCGTAC CCGGCGGCGG	420
GTTCGCGGCG AATATTGCGG GCGGCGGCGG AGACCGGCGG GGGCGGCGAT TGCGGCGGCG	480
CAAGCGAAAC ACAGCGGCGG GTTCGCGGCG GCGGCGGCGG TTGCGGCGAT CACCGGCGAT	540
TCAGCGGCGG CACCGGCGGT AATGTTTATG AACCGGCGTAC CCGGCGGCGG GCGGCGGCGG	600
GCGGCGGCGG GAGGCGGCGG GCGGCGGCGG CGGCGGCGG CAAAAGCGCG GCGGCGGCGG	660
GCGGCGGCGG GAGGCGGCGG GTTCGCGGCG TCGGCGGCGG GCGGCGGCGG CCGGCGGCGG	720
TTGCGGCGGT CAAGCGGTTA GCGGCGGTTG GCGGCGGTTG GCGGCGGCGG CCGGCGGCGG	780
TTGCGGCGGT GTTCGCGGTA GCGGCGGCGG GCGGCGGCGG GTTCGCGGCG TTGCGGCGAT	840
TTGCGGCGGT GCGGCGGCGG GTTCGCGGCG GCGGCGGCGG GTTCGCGGCG GCGGCGGCGG	900
TTGCGGCGGT GCGGCGGCGG	913

(x) INFORMATION FOR SEQ ID NO:17:

1. SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1972 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

GTCATGGGTG	CTGAGCGTGC	TGGCTGCCGT	CGGGCTGGGC	CTGGCCACGG	CGCCGGCCCCA	180
GGCGGCCCG	CCGGCTTGT	CGCAGGACCG	GTTCCGCCAC	TTCGCCCGGC	TGCCCTCGA	240
CCCGTCCGCG	ATGGTCCGCG	AAGTGGCGCC	ACAGGTGGTC	AACATCAACA	CCAAACTGGG	300
CTACAACAAC	GCGTGGGCG	CCGGGACCGG	CATCGTCATC	GATCCCAACG	GTGTCTGCT	360
GACCAACAAC	CACGTGATCG	CGGGCGCCAC	CGACATCAAT	CGCTTCACCG	TGGGTCCCG	420
CCAAACCTAC	GGCGTCCGAT	TGGTCCGGTA	TGACCGCACC	CAGGATGTCTG	CGGTCTGCA	480
GCTCGCGCGT	GCCCGTGGCC	TGCCGTCCGC	GGCGATCGGT	GGCGGCGTCG	CGGTGGTGA	540
GCCCCGTCTC	GCGATGGGCA	ACAGCGGTGG	GCAGGGCGGA	ACGCCCGGTG	CGGTGCTGG	600
CAGGGTGGTC	GCGCTCGGCC	AAACCGTGCA	JGCGTCCGAAT	TCGTGACCG	GTGCCGAAGA	660
GACATTGAAC	GGGTTGATCC	AGTTCGATGC	CGCAATCCAG	CCCGGTGATT	CGGGCGGGCC	720
CGTCTCAAC	GGCCTAGGAC	AGGTGGTCTGG	TATGAACACG	GCCGCGTCCG	ATAACTTCCA	780
GCTGTCCCG	GGTGGGCAGG	GATTCCGCAT	TCCGATCGGG	CAGGCGATGG	CGATCGCGGG	840
CCAAATCCGA	TCCGGTGGGG	GGTCACCAC	CGTTCATATC	GGCGCTACCG	CGTTCCTCGG	900
CTTGGGTGTT	GTGACAACA	ACGGCAACCG	CGCACGAGTC	CAACGCGTGG	TCCGGAAGCG	960
TCCGGCGGCA	AGTCTCGGCA	TCTCCACCGG	CGACGTGATC	ACCGCGGTGG	ACGGCGCTCC	1020
GATCAACTCG	GCACCCCGA	TGGCGGACGC	GCTTAACGGG	CATCATCCCG	GTGACGTCAAT	1080
CTCGGTGAAC	TGGCAACCA	AGTCCGGCGG	CACGCGTACA	GGGAACGTGA	CATTGGCCGA	1140
GGGACCCCG	CCCTGATTTC	TCCGCGGATAC	CACCCGCGCG	CCGGCCCAAT	GGATTGGCGC	1200
CAGCGGTGAT	TGCCCGCTGA	GCCTCCGAGT	TCCGTCTCCC	GTGCGCGTGG	CATTGTGGAA	1260
GCAATGAACG	AGGCAGAAC	CAGCGTTGAG	CACCTCCCG	TGCAGGGCAG	TTACGTGCA	1320
GGCGGTGTGT	TGGAGCATCC	GGATGCCAAG	GACTTCGGCA	GCGCCGCGCG	CCTGCCCGCC	1380
GATCCGACCT	AGGTTAAGCA	CGCCGTCTTC	TACGAGGTGC	TGGTCCGGGC	GTTCTTGCAC	1440
GCCAGCGCGG	CCGTTCCGN	CGATCTGCGT	GGACTCATCG	ATCGCCTCGA	CTACCTGCAG	1500
TGGCTTGGCA	TGCAGTGCAT	CTGTTCCCGG	CGTTCCTACG	ACTCACCGCT	GCGCGACGGC	1560
GGTTACGACA	TTCCGCACTT	CTACAAGGTG	TGCGCCGAAT	TCCGCACCGT	CGACGATTTG	1620
GTCCGCGTGG	TGCACACCGC	TCACCGGCGA	GGTATCCGCA	TCATCACCGA	CCTGGTGAATG	1680
AATCACACCT	CGAGTTCGCA	CGCCTGGTTT	CAGGAGTCCG	GCCGCGACCG	AGACGGACCG	1740
TACGGTGAAT	ATTACGTGTG	GAGCGACACG	AGCGAGCGGT	ACACCGACCG	CCGGATCATC	1800
TCCGTGACA	CCGAAGAGTC	GAACTGGTCA	TTCGATCCGT	TCCGCGGACA	GTTNCTACTG	1860
GCACCGATTC	TT					1872

2 INFORMATION FOR SEQ ID NO:18

2. SEQUENCE CHARACTERISTICS

A. LENGTH: 1482 base pairs

3. TYPE: nucleic acid

C: STRANDEDNESS: single

TOPOLOGY linear

NO	SEQUENCE	DESCRIPTION	REF	NO	NO
1	1	1	1	1	1
2	2	2	2	2	2
3	3	3	3	3	3
4	4	4	4	4	4
5	5	5	5	5	5
6	6	6	6	6	6
7	7	7	7	7	7
8	8	8	8	8	8
9	9	9	9	9	9
10	10	10	10	10	10
11	11	11	11	11	11
12	12	12	12	12	12
13	13	13	13	13	13
14	14	14	14	14	14
15	15	15	15	15	15
16	16	16	16	16	16
17	17	17	17	17	17
18	18	18	18	18	18
19	19	19	19	19	19
20	20	20	20	20	20
21	21	21	21	21	21
22	22	22	22	22	22
23	23	23	23	23	23
24	24	24	24	24	24
25	25	25	25	25	25
26	26	26	26	26	26
27	27	27	27	27	27
28	28	28	28	28	28
29	29	29	29	29	29
30	30	30	30	30	30
31	31	31	31	31	31
32	32	32	32	32	32
33	33	33	33	33	33
34	34	34	34	34	34
35	35	35	35	35	35
36	36	36	36	36	36
37	37	37	37	37	37
38	38	38	38	38	38
39	39	39	39	39	39
40	40	40	40	40	40
41	41	41	41	41	41
42	42	42	42	42	42
43	43	43	43	43	43
44	44	44	44	44	44
45	45	45	45	45	45
46	46	46	46	46	46
47	47	47	47	47	47
48	48	48	48	48	48
49	49	49	49	49	49
50	50	50	50	50	50
51	51	51	51	51	51
52	52	52	52	52	52
53	53	53	53	53	53
54	54	54	54	54	54
55	55	55	55	55	55
56	56	56	56	56	56
57	57	57	57	57	57
58	58	58	58	58	58
59	59	59	59	59	59
60	60	60	60	60	60
61	61	61	61	61	61
62	62	62	62	62	62
63	63	63	63	63	63
64	64	64	64	64	64
65	65	65	65	65	65
66	66	66	66	66	66
67	67	67	67	67	67
68	68	68	68	68	68
69	69	69	69	69	69
70	70	70	70	70	70
71	71	71	71	71	71
72	72	72	72	72	72
73	73	73	73	73	73
74	74	74	74	74	74
75	7				

TTTCTGGGAA	AGCTGATGCG	TAGGAAAGG	TTCTTCGGCT	GAGGCTTACG	GCCTCGGAGC	1
GGTGGTGGCT	CGCCGAGATG	AAGGATGCG	TTGATGGGAC	AAAAAGGCTTG	ACCAAGCTGC	2
AATTAAGGCT	CGGAAACAAAG	TGGAAAGTGG	ACAGGCTTGT	TGTTATTACG	AGTCCGATG	3
TGTAGCTGCG	CGCCAAATCGG	CTTGATGGAA	AGGGGCTATT	CAGCTACAGC	GACGAGGAGG	4
TTCTTGGCTG	TGGGCTAGAA	GGGACACAGC	TGTGGGTGAA	AGTGTTCGAG	GACTGGAGCA	5
ATCTTGGCTG	GATTTCTGAA	TTGTAAATG	CACCGCTGCT	CGATGCTGGC	GCTGGGGTGA	6
CGGAGCTGCT	GTCGGCTGTG	ACGAACTGCG	AAGCGCAAGG	TACCGAAGTG	ATAGACGGAA	7
TTTGGAGGAG	CAAAATCTCG	GGGACGATCG	CGCGGAGCTC	TGTGAAGATG	TTGTATGCTG	8
GGGCGAGAG	TGCAAGGCTG	AGGATCTGCG	AGATTCGCGA	GGAGGCTGCG	TACCGGCTCG	9
CTTGAAGTAA	ATGCGAGCTG	ATATCTGCGC	CGATCGAGCT	TACCGATCTG	TACCGATCTG	10
AAATGCTTAA	ATTGATGATA	TTGATGATA				11

CGGTCGATCA	TGGCGGCGAGC	AACGACGTGG	TCGCGCGCCGA	AAACCTCGCC	CCACCGGCGG	900
AAGGCCCTTAT	TGGACGTGAC	GATCAAGCTG	GCCCGCTCAT	ACCGGGAGGA	CACCAGCTGG	960
AAGAAGAGGT	TGGCGGCGCTC	GGGCTCAAAC	GGAATGTAAC	CGACTTCGTC	AACCACCAGG	1020
AGCGGATAGC	GGCCAAACCG	GGTGAGTTCC	GCGTAGATGC	GCCCCGGCGTG	GTGAGCGCTCG	1080
GCGAACCGTG	CTACCCATTC	GGCGGCGGTG	GCGAACAGCA	CCCGATGACC	GGCTGACAC	1140
GCGCGTATCG	CCAGGCGGAC	CGCAAGATGA	GTCTTCCC	TGCCAGGCGG	GGCCCCAAAA	1200
CACGACGTTA	TCGCGGGCGG	TGATGAAATC	CAGGGTGCCC	AGATGTGCGA	TGGTGTGCGG	1260
TTTGAGGCCA	CGAGCATGCT	CAAAGTCGAA	CTCTTCCAAC	GACTTCCGAA	CCGGGAAGCG	1320
GGCGGCGCGG	ATGCGGCCCT	CACCACCATG	GGATCCCGG	GCTGACACTT	CCCGCTGCAG	1380
GCAGGCGGCC	AGGTATTCTT	CGTGGCTCCA	GTTCTCGGCG	CGGGCGCGAT	CGGCCAGCGG	1440
GGACACTGAC	TCACGCAGGG	TGGGAGCTTT	CAATGCTCTT	GT		1482

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 876 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GAATTCGGCA	CGAGCGCGCG	ATAGCTTCTG	GGCGCGGCGC	GACCAGATGG	CTCGAGGGTT	60
CGTGCTGGG	GCCACCGCGG	GCGGCACCA	CCTGACCGGT	GAGGGCGTGC	AACACGCCGA	120
CGGTCACTCG	TTGCTGCTGG	ACGCCACCAA	CCCGGCGGTG	GTTGCCTACG	ACCGGCGCTT	180
CGCCTACGAA	ATCGGCTACA	TCGNGGAAAG	CGGACTGGCC	AGGATGTGGG	GGGAGAACCC	240
GSAGAACATC	TTCTTCTACA	TCACCGTCTA	CAACGAGCCG	TACGTGCAGC	CGCGGAGGCC	300
GSAGAACTTC	GATCCCCBAGG	GCCTGCTGGG	GGGTATCTAC	CGNTATCAGG	CGGCCACCGA	360
GCAACGCACC	AACAAGGNGC	AGATCCTGGG	CTCGGGGGTA	GCGATGCCCC	CGGCGCTGCG	420
GCGAGCACAG	ATGCTGCGCG	CCGAGTGGGA	TGTGCGCGGC	GACGTGTGGT	CGGTGACCAG	480
TTGGGCGCAG	CTAAACCTGG	ACGGGTGGGT	CATCGAGACT	GAGAACTTCC	GCCACCGCGA	540
TCGGCGCGCG	GGCGTGCCCT	ACGTGACGAG	AGCGCTGGAG	AATGCTCGGG	GCCCGGTGAT	600
TCGGGTGTCT	GACTGGATGC	GGCGGCTGCG	CGAGCAGATC	CGACCTGGG	TGCGGGGAC	660
ATACCTCAGC	TTGGGCGAGG	ACGGGTTCGG	TTTTTCCGAC	ACTCGGCGCG	CGGCTCGTGG	720
TTACTTCCAC	ACCGACGCTG	AATCCGAGGT	TGGTCCGCGT	TTTGGGAGGG	GTTGGCGCGG	780
TGGACGGGTG	AATATCGACC	CATTCGGTGC	CGGTCTGGG	TCGCGCGCGC	AGTTACCGCG	840
ATTTCAGCAA	GCTGGGGGGT	TGCGCGCGAN	TAAGTT			876

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1071 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATCGCTGCGG	GCTGCAGGAA	TTCTTCTACA	GAAGCAAAAT	TCATCTGCTT	AATCCAGGAA	60
AGAGTTTATA	AGCAATTTAC	ACCGGACCAA	CAAGATGTGC	AGATCTGCGA	TTATTTCCAG	120
AGCGAATATC	TCGGCGAGCT	TCGAGGAAAT	TTTTCAGCGA			
ATATCAATTA	TCGCTGCTGG					

TTGATGGCAA	CCCTGGTGGC	GGTTGCCGAT	CGGGCCGGGG	CCAACCTGTT	CGAGCTAGAG	480
AACCTTCGTG	CACGTGAAGT	GGATGTGGCG	CCGGCCGCAT	CAGGCGCCCC	GCACGCTGCC	540
GGGGGCCGCG	TCTAGATCCC	TGGGGGGGAT	CAGCGAGTGG	TCCCSTTCGC	CCGCGCGTCT	600
TCCAGCCAGG	CCTTGGTGGC	CCCGGGGTGG	TGAGTACCAA	TCCAGGCCAC	CCCGACCTCC	660
CGNAAAAGT	CGATGTCCCT	GTAATCATCG	ACGTTCCAGG	AGTACACCGC	CCGGCCCTGA	720
GCTGCCGAGC	GGTCAACGAG	TTGCCGATAT	TCCTTTAACC	CAGGCAGTGA	GGGTCCACAG	780
GCGGTTGGCC	CGACCGCCGT	GGCCGCACTG	CTGGTCAGGT	ATCGGGGGGT	CTTGGCGGAG	840
AACAACGTG	GCAGGAGGGG	TGGAGCCCGC	CGGATCCGCA	GACCGGGGGG	GCGAAAACGA	900
CATCAACACC	GCACGGGATC	GATCTGCGGA	GGGGGGTGCG	GGATATACCG	ACCGGTGTAG	960
GAGCGCCAGC	AGTTGTTTTT	CCACCAGCGA	AGCGTTTTTC	GGTCATCGGN	GGCNNTTAAG	1020
T						1021

(2) INFORMATION FOR SEQ ID NO:21:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 321 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CGTGGCGAGC	AACGGAAGAA	CACAACCATG	AAGATGGTGA	AATCGATCGC	CGCAGGTCTG	60
ACCGGCGCGG	CTGCAATCGG	CGCGGCTGCG	GCCGGTGTGA	CTTCGATCAT	GGCTGGCGGN	120
CGCGTCTGAT	ACCAGATGCA	GCCGGTCTGT	TTGGCGCGCG	CCTGCGCGTT	GGACCGGNA	180
TCCGCGCGCTG	ANGTCCCGAC	CGCGCGCGAG	TGGACCGAGC	TGCTCAACAG	NCTCGNCGAT	240
CCCAACGTGT	CGTTTGNGAA	CAAGGGNAGT	CTGGTGGAGG	GNGGNATCGG	NNGNANCGAG	300
GNGNGNATCG	GNGNANCGAG	A				321

(2) INFORMATION FOR SEQ ID NO:22:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 373 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TCTATCGGCT	TCCGGTTGGG	GACCGGTATT	GGNGCGGGGT	CTTTAACCGG	CTCGGCGAGG	60
CTATCGACCG	GTCGGGAGAG	GTGACTCCCG	ATAGTGGGCG	CTCGCTGGAG	CTCGAGGCGG	120
CTTGGGTGCT	GNACCGGCAA	GGGTGAAGG	AGCGGTTGNA	GACCGGGATC	AAGGTGATTG	180
AGGAGATGAC	CCCGATCGCG	CGCGGCGAGT	GGGAGGTGAT	CTTCGGGGAG	CGCAAGACCG	240
CTTAAACCGG	CGGTCTGTGT	CGGACACCAT	CTTAAACCGA	CGCGGAAGAA	CTGGGAGTCT	300
CTTGGATCCC	AAGAAGTAGG	TCCGCTTGTG	TATAGTTTGG	GCATCGGGCA	AGAAGGGGAA	360
CTTAACTCGG	CGG					373

(2) INFORMATION FOR SEQ ID NO:23:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 373 base pairs
- (B) TYPE:

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GTGACGCCGT GATGGGATTC CTGGGCGGGG CCGGTCCGCT GGUGGTGGTG GATCAGCAAC	60
TGGTTACCCG GGTGCCGCAA GGCTGGTGGT TTGCTCAGGC AGCCGCTGTG CCGGTGGTGT	120
TCCTTGACGGC CTGGTACGGG TTGGCCGATT TAGCCGAGAT CAAGGCGGGC GAATCGGTGC	180
TGATCCATGC CGGTACCGGC GGTGTGGGCA TGGCGGCTGT GCAGCTGGCT CGCCAGTGGG	240
GCSTGGAGGT TTTCTGCACC GCCAGCCGTG GNAAGTGGGA CACGCTGGCG GCCATNGNGT	300
TTGACGACGA NCCATATCGG NGATTCCCN CACATNCGAAG TTCCGANGGA GA	352

(2) INFORMATION FOR SEQ ID NO:24:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 726 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GAAATCCGGG TTCATTCCGT TCGACCAGCG SCTGGCGGATA ATCGACGAAG TGATCAAGCC	60
GCGGTTCGCG GCGCTCATGG GTCACAGCGA GTAATCAGCA AGTCTCTGGS TATATCCGAC	120
CTAGGCTCCA GTTGCTTGCC AGATCGCTTT CGTACCGTCA TCGCATGTAC CCGTTTCGCGT	180
GCCGACGCT CATGCTGGCG GCGTGCATCC TGGCCACGGG TGTGGCGGGT CTCGGGGTCC	240
GCGGSCAGTC CGCAGCCCAA ACCGCGCCGG TGGCCGACTA CTACTGGTGC CCGGGGCAGC	300
CTTTGACCCG CGCATGGGGG CCCAACTGGG ATCCCTACAC CTGCCATGAC GACTTCCACC	360
GCGACAGCGA CGGCCCCGAC CACAGCCGGG ACTACCCCGG ACCCATCCTC GAAGGTCCCG	420
TGCTTGACGA TCCCGGTGCT GCGCCGCGCG TCCCGGCTGC CGGTGGCGGG GCATAGCCT	480
CGTTGACCGG GCGGCTCAG CGAATACCG TATAAACCGG GCGGTGCCCG CCGCAAGCTA	540
CGACCCCGCG CGGGGCGAGT TTACGCTCCG GTGCGGATG ATCGCCCGCG CCGATGACAG	600
AAAATAGGGG ACGGTTTTGG CAACCGCTT GAGGACGCTT GAAGGGAACC TGTATGAAC	660
GGGACAGCGG CTTCCACCAT CGACATCGAC AAGGTTGTTA CCGGCACACC CGTTCCCGCG	720

(2) INFORMATION FOR SEQ ID NO:25:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 680 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:25:

AGGATGAGG AGGAACGTGG GCGGACGAG CCGCTATGCG TTTATGCAAG CGACCGGGAT	60
GCTGCGCGAG CATATCCGAG CATGCTGGT GCGCACTGAG CGACTTTTGG ACCAGCCCGG	120
CTGCTGAGG GCGGCGCGGT GAAGTCAAT GCGGCGGGCT TGTGCACTG ATGAACCGGA	180
ATAGGGAAGA ATAGGCGGGT GATTTGGGAG TTGAATGTGG GTATGCGTG GAAATCCGAT	240
GCGGCGGGCAT GCTCGGCGCG GACGAGGTC GCGGAGGCGG GCGAGGCTGA ATCTGGAGGG	300
AGCACTCAAT GCGGCGCGAT AAGGCGCGGA CCGGCGACCG TCGTTTGGAA GCAACTAAGG	360
AGGCGCGCGG CATTTGCGTG CGAGTATTA TTGAGGCTGG CGTATGATG CTCGTCCGAG	420
TGACAGCGGA CGAAGCGCGG CGACTGCTT AGGAAGTCAA AGGCTTATG AGGATATG	480
AGGCAATGCG CATATGCTGG GCTTTATG	540

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 160 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

AACGGAGGCG CCGGGGGTTT TGGCGGGGCC GGGGCGGTCG GCGGCAACGG CCGGGCCGGC 60
 GGTACCGCCG GGTGTTCGG TGTCCGGGGG GCCGGTGGGG CCGGAGGCAA CCGCATCGCC 120
 GGTGTCACGG GTACGTCGGC CAGCACACCG GGTGGATCCG 160

(2) INFORMATION FOR SEQ ID NO:27:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 172 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GACACCGATA CGATGGTGAT GTACGCCAAC GTTGTGACA CGCTCGAGGC GTTCACGATE 60
 CAGCCGCACAC CCGACGGCGT GACCATCGGC GATCGGGCCC CGTTCGCGGA GGCGGCTGCC 120
 AAGGCGATCG GAATCGACAA GCTGCGGGTA ATTACATCCG GAATGGACCC CGTCGTCGCT 180
 GAACGCGAAC AGTGGGACGA CCGCAACAAC ACGTTGGCGT TGGCGCCCGG TGTCGTTGTC 240
 CCTACGAGC GCAACGTACA GACCAACGCT CG 272

(2) INFORMATION FOR SEQ ID NO:28:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 317 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCAGCGGGTG GTTCTGGGAC TATCTGGGCA TGGTACGCA GCGGACGTC GCGGAGGTGA 60
 AGCGGATCGA GCAGAGGAT TGGTGGGCG GTTCATGCG GTACGTGGCC GGTATCACCG 120
 GCGAGGAGGT GAATCTGGT GAACCGGCGC GGTGATGCG GTGAGCGCG GCGACGATCC 180
 GTTCGGATCT GCGGTGGTTC GAGACGCTCT ATCTGTACA TGGGTTGGCC GCTGGGTCGC 240
 GGAATCTGAT GCGGAGGATC AGAAGCGCT DAAAGATTA GTTCTGAC ACTGGCTTCG 300
 GCGCTTGCTT GCGCGGG 317

(2) INFORMATION FOR SEQ ID NO:29:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 181 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single

GATCGTGGAG CTGTGGATGA ACAGCGTTGC CGGACGCGCG GCGGCCAGCA CCGCGGTGTA 60
 GCAGCGCCGG ACCACCTCGC CGGTGGGCAG CATGGTGATG ACCACGTCGG CCTCGGCCAC 120
 CGCTTCGGGC GCGCTACGAA ACACCGCGAC ACCGTGCGCG GCGGCGCCGG ACGCCGCCGT 180
 GG 182

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 308 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GATCGGGAAG TTTGGTGAAC AGGTGGTGA CCGGAAAGTC TGGGCGCCTG CGAAGCGGGT 60
 CGGCGTTTAC GAGGCGAAGA CAGCCTGCTC CGAGCTGCTG CCGCTGCTCT ACGGCGGGCA 120
 GAGTTTGAGA TTGCGCGCGG CCGCGAGCGG GTAGCAAAGC TTGTGCGGCT GCATCCTCAT 180
 GAGACTCGGC GGTAGGCAAT TGACCATGGC GTGTACCGCG TGCCCGACGA TTGGACGCT 240
 CGTTGTGAG ACCACGTGCT CGAAGCTTT CACCGTGAA GCGCTACCTC ATCGACACCC 300
 AGGTTTGG 308

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 267 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CGGACGACCA GCAACTCAGC TGGATGATG TCGGCGAGCG CATGAGGAGC GGAGAGAATC 60
 CGGCGGAAGC TCGCGCGCGG CAACTGCTCA TAGTGACCGG CGGTAGAGGG CTCGCGCGAT 120
 CGGACGCGAC TATTCTGCTG TCGCGCTGGC CGGTAAGAGC GGGTAAAAGA ATGTGAGGGG 180
 ACAGGATGAG CAATCAGATC TACCGAGTGA TCGAGATCGT CGGACCTCG CCGGACGGCG 240
 TCGACGCGGC AATCCAGGGC GGTCTGC 267

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1544 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CTGCTGCGGA AAGAATGTGA GGGGACGGA TGACCAATCA CAGCTACCGA GTGATCGAGA 60
 TGTGCGGAC CTGCGCGGA GCGCTGCAAT GCGCAATCCA GGGGCTCTG GCGCGAGCTG 120
 TGAAGACAT GCGCTGCTG TACTGCTTGA AATACAGTC AATTCAGATC TACTGCTG 180
 AATGAGCTT TCGCAATGTA TACTGCTTGA TACTGCTTGA TACTGCTTGA TACTGCTTGA 240

GCTCGACCTG	TGTGGGCTGC	AGCCGGACGA	AGCGGTGCTC	GACGTCGGCT	GCGGCTCGGG	480
GCGGATGGCG	TTGCCGCTCA	CCGGCTATCT	GAACAGCGAG	GGACGCTACG	CCGGCTTCGA	540
TATCTCAG	AAAGCCATCG	CGTGGTGCCA	GGAGCACATC	ACCTCGGCGC	ACCCCAACTT	600
CCAGTTCGAG	GTCTCCGACA	TCTACAACTC	GCTGTACAAC	CCGAAAGGGA	AATACCAGTC	660
ACTAGACTTT	CGCTTTCCAT	ATCCGGATGC	GTCGTTGAT	GTGGTGTTC	TTACCTCGGT	720
GTTCAACCCAC	ATGTTTCCGC	CGGACGTGGA	GCACTATCTG	GACGAGATCT	CCCGCGTGCT	780
GAAGCCCGGC	GGACGATGCC	TGTGCACGTA	CTTCTTGCTC	AATGACGAGT	CGTTAGCCCA	840
CATCGCGGAA	GGAAAGAGTG	CGCACAATT	CCAGCATGAG	GGACCGGGTT	ATCGGACAAT	900
CCACAAGAAG	CGGCCCGAAG	AAGCAATCGG	CTTGCCGGAG	ACCTTCGTCA	GGGATGTCTA	960
TGGCAACTTC	GGCCTCGCCG	TGCACGAACC	ATTGCACTAC	GGCTCATGGA	GTGGCCGGGA	1020
ACCACGCCTA	AGCTTCCAGG	ACATCGTCAT	CGCGACCAAA	ACCGCGAGCT	AGGTCCGCAT	1080
CCGGGAAGCA	TCGCGACACC	GTGGCGCCGA	GCGCCGCTGC	CGGCAGGCCG	ATTAGCCGGG	1140
CAGATTAGCC	CGCCCGGGCT	CCCGGCTCCG	AGTACGGCGC	CCCGAATGGC	GTCAACCGGT	1200
GGTAACCACG	CTTGCGCGCC	TGGGCGCGCG	CCTGCCGGAT	CAGGTGGTAG	ATGCCGACAA	1260
AGCCTGCGTG	ATCGGTGATC	ACCAACGGTG	ACAGCAGCCG	GTTGTGCACC	AGCGCGAACG	1320
CCACCCCGGT	CTCCGGGTCT	GTCCAGCCGA	TCGAGCCGCC	CAAGCCCAAC	TGACCAAAAC	1380
CCGGCATCAC	GTGCGCGATC	GGCATACCGT	GATAGCCAAG	ATGAAAATTT	AAGGGCACCA	1440
ATAGATTTTG	ATCCGCGAGA	ACTTCCGCTC	GGTTCCGGGT	CAGGCCCGTG	ACCAGCTCCC	1500
CCGACAGAA	CCGTATGCCG	TGATCTCCG	CTCCTGCCG			1560

(2) INFORMATION FOR SEQ ID NO:33:

(A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 351 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CTCCAGGGTG	GCSTGGATGA	GCSTCACCBC	GGGSCAGGCC	GAGCTGACCG	CCGCCCCAGGT	60
CCGGGTTGCT	GCGGCGGCTT	ACGAGACGGC	GTATGGGCTG	ACGGTGCCCC	CCCGCGGTGAT	120
CCCGGAGAAC	CGTGCTGAAC	TGATGATTCT	GATAGCGACC	AACCTGTTGG	GGCAAAACAC	180
CCCGCGGATC	GCGGTCAACG	AGGCCGAATA	CCSCGAGATG	TGGGCCCAAG	ACGCCGCCGC	240
GATGTTTGGC	TACGCGCGCG	CGACCGCGAC	GGCGACGGCG	ACGTTGCTGC	CGTTCGAGGA	300
CCCGCCGGAG	ATGACCAGCG	CGGTTGGGCT	CCTCGAGCAG	CCCGCCCGCG	TGGAGGAGGC	360
CTCCGACACC	GCCCGCGCGA	ACCASTTGAT	GAACAATGTC	CCCGAGGCCG	TGAAACAGTT	420
GGCCGAGCCC	ACGCGAGGCA	CCAGGCTTC	TTCCAAGCTG	TGTGGCCTGT	GGAAGACGGT	480
CTCCCGCAT	CGGTGCGCGA	TCAGCAACAT	GCTGTGATG	CCCAACAACC	ACATGTGAT	540
GACCAACTCG	GCTGTGTCGA	TGACCAACAC	CTTGAGCTCG	ATGTTGAAGG	GCTTTGCTCG	600
CCCGCGCGCG	GCCGAGGCGG	TGCAAACTCG	CCCGCAAAAC	GGGTTCCGGG	CGATGACCTG	660
CTTGGCGAGC	TGGCTGGGTT	TTTGGGCTCT	CCCGGCTGCG	TTGGGCGCGA	ACTTGGGTCC	720
CCCGGCTCC	GTACGCTATG	ATTAAGGCGA	TCCCGGAAAA	TATGCAAAAT	CTGGTGGGCG	780
GAAGGCTGCT	CCCGGCTAAG	CTTACGCTCG	ATTTTCTGGA	TGGGTGAAC	TTGCTCAACG	840
GAAGAGTTA	C					900

(2) INFORMATION FOR SEQ ID NO:34:

(A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 351 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

GATCGATCGG	GCGGAAATTT	GGACCAGATT	CGCCTCGGGC	GATAACCCAA	TCAATCGAAC	60
CTAGATTTAT	TCCGTCCAGG	GGCCCGAGTA	ATGGCTCGCA	GGAGAGGAAC	CTTACTGCTG	120
CGGGCACCTG	TGCTAGGTCC	TGGATACGGC	GGGAGGCGTC	GACATTTTCC	ACCGACACCC	180
CCATCCAAAC	GTTCGAGGGC	CACTCCAGCT	TGTGAGCGAG	GCGAGCGAGT	CGCAGGCTGC	240
GCTTGGTCAA	GATC					254

(2) INFORMATION FOR SEQ ID NO:35:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1227 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GATCCTGACC	GAAGCCGCTG	CGCCCAAGGC	GAAGTCGCTG	TTGGACCCAGG	AGGGACCGGA	60
CGATCTGGGC	CTGCCGATCG	CGGTTCAAGC	GGGGGGGTGC	GCTGGATTGC	GCTATAACCT	120
TTTCTTCGAC	GACCCGACGC	TGGATGGTGA	CCAAACCGCG	GAGTTGGGTG	GTGTCCAGTT	180
GATCGTGGAC	CGGATGACCG	CGCCGTATGT	GGAAGGCGCG	TGGATCGATT	TGCTCGACAC	240
TATTGAGAAG	CAAGGTTCA	CATCGACAAT	CCCAACGCCA	CGGGCTCCTG	CGCGTGCGGG	300
GATTCGTTCA	ACTGATAAAA	CGCTAGTACG	ACCCCGCGGT	GCGCAACAGG	TACGAGCACA	360
CCAAGACCTG	ACCGCGCTGG	AAAAGCAACT	GAGCGATGCC	TTGCACCTGA	CGCGCTGGCG	420
GGCCGCGCGG	GGCAGGTGTC	ACCTGCATGG	TGAACAGCAC	CTGGGCGCTG	TATTGCGACC	480
AGTACACGAT	TTTGTGATC	GAGGTCACCT	CGACCTGGGA	GAACTGCTTG	CGGAACCGCT	540
CGCTGCTCAG	CTTGGCCAA	GCCTGATCGG	AGCGCTTGTG	GCGCACGCGG	TGCTGGATAC	600
CGCACAGCGC	ATTGCGAAGC	ATGGTGTCCA	CATCGCGGTT	CTCCAGCGCG	TTGAGGTATC	660
CCTGAATCGC	GGTTTTGGCC	GGTCCCTCCG	AGAATGTGCC	TGCCGTGTTG	GCTCCGTTGG	720
TCCCGACCCC	GTATATGATC	GGCGCCGTCG	TAGCCGACAC	CAGCGCGAGG	GCTACCACAA	780
TGCGGATCAG	CAGCGGCTTG	TGCGGTGCGT	TGCGGTAGGA	GACCTGCGGT	GGCACCGCGG	840
GATATCGCGC	GGCGCGGAGC	GGCGGCTCGT	CTGCCGGTTC	CGGGGCGAAG	CCCGGTTCCG	900
CGCGCCCGAG	GTGCTGGGGG	TAGTCCAGGG	CTTGGGGTTC	GTGGGATGAG	GGCTCGGGGT	960
ACCGCCCGCG	TCCGTGGGTG	CGACACCGCG	GCTTCCGCGA	GTGGGAGCGG	GCCATTGTGG	1020
CTCTCTTAGG	GTGGTGGAGC	GGACCAGGTC	CTAGGGCGAC	AACTGCGCGT	TGCTCAGGCC	1080
GGCAGCATCG	GCAATCAGGT	GAGCTCCCTA	GGCAGGCTAG	TGCAACAGCT	GCCCTCAGCT	1140
CTCAACCCGA	CGGGGCGGGT	CGCGGCGCGG	ATAATGTTGA	AAGACTAGGC	AACCTTAGGA	1200
ACGAGCGACG	GAGATTTTGT	GACGATC				1227

(2) INFORMATION FOR SEQ ID NO:36:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 181 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

TTTCTTTTGG	TTGATCTGGT	TTTCTTTTGA	ACTTCAACCG	TTGATCTGGT	TTTCTTTTGG	60
TTTCTTTTGG	TTGATCTGGT	TTTCTTTTGA	ACTTCAACCG	TTGATCTGGT	TTTCTTTTGG	120

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 290 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GCGGTGTCGG	CGGATCCGGC	GGGTGGTTGA	ACGGCAACGG	CGGTGTCGGC	GGCCGGGGCG	60
GCGACGGCGT	CTTTGCCGGT	GCCGGCGGCC	AGGGCGGCCT	CGGTGGGCAG	GGCGGCAATG	120
GCGGCGGGTC	CACCGCGGGC	AACGGCGGTC	TTGGCGGGCG	GGGCGGTGGC	GGAGGCAACG	180
CCCCGGACGG	CGGCTTCGGT	GGCAACGGCG	GTAAGGGTGG	CCAGGGCGGN	ATTGGCGGGC	240
GCACTCAGAG	CGCGACCGGC	CTCGGNGGTG	ACGGCGGTGA	CGGCGGTGAC		290

(2) INFORMATION FOR SEQ ID NO:38:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 34 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GATCCAGTGG	CATGGNGGCT	CTCACTGGAA	GCAT	34
------------	------------	------------	------	----

(2) INFORMATION FOR SEQ ID NO:39:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 155 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GATCGGTGCT	CGTCCCCCCC	TTGCGGCGCA	CGCCACCGGT	CCGACCGTTA	CCGAACAAGC	60
TGGGCTGGTC	CCCAGCACCC	CGGGCAUCCG	CGACCGCGGA	CTCGAAGCAT	GGCAGCGTCS	120
TATCGGCAAC	ATTGCGGCGG	CTTCGAGTGG	TACCG			155

(2) INFORMATION FOR SEQ ID NO:40:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 57 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

- (1) SEQUENCE CHARACTERISTICS.
 (A) LENGTH: 132 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GATCCACCGC GGGTGCAGAC GGTGCCCCGG GCGCCACCCC GACCAGCGGC GGCAACGGCG	60
GCACCGGCGG CAACGGGCGG AACGCCACCG TCGTCGGNGG GGCCGGCGGG GCCGGCGGCA	120
AGGGCGGCAA CG	132

(2) INFORMATION FOR SEQ ID NO:42:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 132 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GATCGGGGGG CGGNACGGNC GGGGACGGCG GCAAGGGCGG NAACGGGGGG GCCGNAGCCA	60
CCNGCCAAGA ATCCTCCGNG TCNCCAATG GCGGAATGG CGGACAGGGG GGCAACGGCG	120
GCANCGGCGG CA	132

(2) INFORMATION FOR SEQ ID NO:43:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 702 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

CGGCACGAGG ATCGGTACCC TGGGGCATCG GCAGCTGCGG ATTCGCGCGG TTTCGCCACC	60
CGAGGAAAGC CGCTACCAGA TGGCGCTGCC GAAGTAGGGG GATCGGTTCG CGATGCCGGC	120
ATGAACGGGC GGCATCAAAT TAGTGCAGGA ACCCTTCAGT TTAGCGACGA TAATGGCTAT	180
AGCACTAAGG AGGATGATCC CATATGACCG AGTCCGAGAC CTTGACGGTG GATCAGCAAG	240
AGATTTTGA AAGGGGCAAC TAGGTGAGG TCCGATGGC GGACGCACCG ACTGATGTCC	300
GCATCACACC GTGCGAATC ATGGGCGNTA AAAACGCCCC TCAACAGNTG GTNTTGTCCG	360
CGGACAACAT GCGGGAATAC TTGGGGGTCG TTGCAAGA GCGGZAGCGT CTGGCGACCT	420
CGCTGCGCAA CGCGGCCAAG GNGTATGCGG AGGTTGATGA GAGGCTGCG ACCGCGCTGG	480
ACAACGACCG CGAAGGAACT GTTCAGGTCG AATCGGCCGG GCGCGTCGGA GGGGACAGTT	540
CGGCGGAACT AATCGATACG CGGAGGGTGG CGACGGCCTG TGAACCCAAC TTCATGGATC	600
TCAAGAAGC GGCAAGGAAG CTCGAAACCG GCGACCAAGG CGCATCCCTC GCGCACTGNG	660
GGGATGGGTG GAACACTTNC ACCCTGACCG TGAAGGGCGA CG	702

(2) INFORMATION FOR SEQ ID NO:44:

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GAAGCCGCAG CGCTGTCGGG CGACGTGGCG GTCAAAGCGG CATCGCTCGG TGGCGGTGGA	60
GGCGCGCGGG TGCCGTCGGC GCCGTGCGGA TCCGCGATCG GGGCGCCCGA ATCGGTGCGG	120
CCCCTGCGCG CTGGTGACAT TGCCGCTTA GCCCAGGAA GGGCGCGCG CGGCGCGCG	180
CTGGCGCGCG GTGGCATGGG AATGCCGATG GGTGCCGCG ATCAGGGACA AGGGGCGCG	240
AAGTCCAAGG GTTCTCAGCA GGAAGACGAG GCGCTCTACA CCGAGGATCC TCGTGCCG	298

(2) INFORMATION FOR SEQ ID NO:45:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1058 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

X1 SEQUENCE DESCRIPTION: SEQ ID NO:45:

CGGCGCGAGG ATCGAATCGG GTCGCGCGGA GCACAGCGTC GCACTGCACC ACTGGAGGAG	60
CCATGACCTA CTCGCGCGGT AACCCCGGAT ACCCGCAAGC GCAGCCCGCA GCCTCCTACG	120
GAGCGCTCAC ACCCTCGTTC GCCCACGCGG ATGAGGGTGC GAGCAAGCTA CCGATGTACC	180
TGAACATCGC GGTGGCAGTG CTCGGTCTGG CTGCGTACTT CGCCAGCTTC GGCCCAATGT	240
TCACCGTCAG TACCGAATC GGGGGGCGTG ATGGCGCAGT GTCCGGTGAC ACTGGGCTGC	300
CGGTGCGGGT GGCTCTGCTG GCTGCGCTGC TTGCGCGGGT GGTCTGTGGT CCGAAGGCCA	360
AGAGCGCATGT GACCGTAGTT GCGGTCTCTG GGTACTCGG CATTATTTCTG ATGGTCTCGG	420
CGACGTTTAA CAAGCGCGAG CCTATTTCGA CCGGTTGGGG ATTGTGGGTT GTGTTGCTT	480
TCATCGTGT CCAGCGCGTT GCGGCGAGTC TGCGGCTCTT GGTGGAGACC GGCGCTATCA	540
CGGCGCGCGG CCGCGCGGCC AAGTTGAGCC CSTATGGACA GTACGGCGCG TACGGGAGT	600
ACGGGCGAGTA CGGCGTGGAG CCGGCTGGGT ACTACGGTCA GAGGGTGTGT CAGCGAGCGG	660
CGGCGAGTCA GTGCGCGCGG CCGGCGAGT CTCGCGAGCG TCCCGGATAT GGTGCGAGT	720
ACGGCGGCTA TTGTTGAGT GCGAGCGAAT CGGCGAGTGG ATACAATGCT CAGCGCGCTG	780
CGGCGCGCGG GCGCGAGTGC GGTGCGCAAC AATCGCACCA GGGCGATGCG ACGCGAGCTA	840
CGGCGTTTCC GAGTTTCAGG GCACGACGCG CGGTCACTGC CGGGAAGGGG TCGCAGGCTG	900
GTTCGGCTCC AGTCAACTAT TCAAAACCCCA GCGGGGGCGA GCAGTCTGCG TCGCGCGCGG	960
GGGCGCGCGT CTAACCGGGC GTTCCCGCGT CCGGTGCGCG GTGTGCGCGA AGAGTGAACA	1020
GGTGTGCGG AAGCGCGGAC GATCCTCGTG CGGAATTC	1058

(2) INFORMATION FOR SEQ ID NO:46:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1058 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

X1 SEQUENCE DESCRIPTION: SEQ ID NO:46:

CGGCGCGAGG ATCGAATCGG GTCGCGCGGA GCACAGCGTC GCACTGCACC ACTGGAGGAG
 CGGCGCGAGG ATCGAATCGG GTCGCGCGGA GCACAGCGTC GCACTGCACC ACTGGAGGAG

(2) INFORMATION FOR SEQ ID NO:47:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 170 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

CGGTCGCGAT GATGGCGTTG TCGAACGTGA CCGATTCTGT ACCGCCGTCG TTGAGATCAA	60
CCAACAACGT GTTGGCSTCG GCAAATGTGC CGNACCCGTG GATCTCGGTG ATCTTGTTCCT	120
TCTTCATCAG GAAGTGCACA CCGGCCACCC TGCCTCCGGN TACCTTTCGG	170

(2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 127 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

GATCCGGCGG CACGGGUGGT GCGGGCGGCA GCACCGCTGG CGCTGGCGGC AACGGCGGGG	50
CGGGGGGTGG CCGCGGAACC GGTGGGTTCG TCTTGGGCAA CGGCGGTGCC GGCGGGCACC	120
GGGCCGT	127

(2) INFORMATION FOR SEQ ID NO:49:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 91 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

AGCGGCAAG GCGGGACCG GCGGACCG GCGGGGCGCA AGCGGGCA	5
AGCGGCTCG GCGGCAACS	91

(2) INFORMATION FOR SEQ ID NO:50:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 149 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION:

(2) INFORMATION FOR SEQ ID NO:51:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 355 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CGGCACGAGA TCACACCTAC CGAGTGATCG AGATCGTCGG GACCTCGCCC GACGGTGTCTG	60
ACGCGGNAAT CCAGGGCGGT CTGGCCCGAG CTGCGCAGAC CATGCGCGCG CTGGACTGGT	120
TCGAAGTACA GTCAATTGGA GGCCACCTGG TCGACGGAGC GGTGCGCGAC TTCCAGGTGA	180
CTATGAAAGT CCGCTTCCGC CTGGAGGATT CCTGAACCTT CAAGCGCGGC CGATAACTGA	240
GGTGCATCAT TAAGCGACTT TTCCAGAACA TCCTGACGCG CTCGAAACCC GGTTCAGCCG	300
ACGGTGGCTC CCGCGACCCG CTGCTTCCAA AATCCCTGCG ACAATTCGTC GCGCG	355

2. INFORMATION FOR SEQ ID NO:52:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 999 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

x1 SEQUENCE DESCRIPTION: SEQ ID NO:52:

ATGCATCACC ATCACCATCA CATGCATCAG GTGGACCCCA ACTTGACAGC TCGCAAGGGA	60
CGATTGGCGG CACTGGCTAT CGCGCGGATG GCCAGCGCCA GCCTGGTGAC CGTTGCGGTG	120
CCCGCGACCG CCAACSCCGA TCGGAGCCCA GCGCCCGCGG TACCCACAAC GGCCGCGCTG	180
TCGCGCTCGA CCGCTGCGAGC GGCACCCGCA CCGCGCGACG CTGTTCCGCC CCCACGACCG	240
TCGCGCGCCA ACACGCGGAA TGCGCAGGCG GCGGATCCCA ACACAGCACC TCGCGCGGCG	300
ACCGGGAACG CACCGCGCGC ACTTGTGATT GCCCAAAAGC CACCCCAACC TGTCGGGATG	360
GACAACCGCG TTGGAGGATT CAGCTTCGCG CTGCTGTGTC GCTGGGTGGA GTCTGACGCG	420
GCGGACITCG ACTACGGTTC AGCACTCCTC AGCAAAACCA CCGGGGACCC GCCATTTCCC	480
GGACAGCGCG CCGCGGTGGC CAATGACACC CGTATCGTGC TCGGCGCGCT AGACCAAAAG	540
CTTTACGCCA GCGCGGAAGC CACCGACTCC AAGGCGCGCG TCGCGTTGGG CTCGGACATG	600
GGTGAGTTCT ATATGCCCTA CCGCGGACCG CGGATTAAGC AGGAAACCGT CTCGCTCGAC	660
TCGAACCGGG TGTGTGGAAG CGCGCTGATC TACCAAGTCA ACTTCAGCCA TCCGAGTAAG	720
TCGAACCGGG AGATCTGGAC GGGGTAATC GGTGCGCGCG TCGGAAACCG ACCTGACGCG	780
GGGCGCGCTC AGCGCTGGTT TGTGTGATCG CTGGGACCGC TCAACAAACG GGTGACAAG	840
GGGCGCGGCA AGGCGCTGGC CGAATGATC CGGCTTTTGG TCGGCGCGCG GCGCGCGCGC	900
GCACCGGCTC CTGCAGAGCG TGTGTGCGCG CCGCGCGCGG TCGGCGGAGT CGCTCTTACC	960
TCGAACGACG CGACACCGCA GCGGATTTTA TCGGCTTGA	999

3. INFORMATION FOR SEQ ID NO:53:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 111 base pairs
 (B) TYPE: linear

Met His His His His His Met His Gln Val Asp Pro Asn Leu Thr
1 5 10 15
Arg Arg Lys Gly Arg Leu Ala Ala Leu Ala Ile Ala Ala Met Ala Ser
20 25 30
Ala Ser Leu Val Thr Val Ala Val Pro Ala Thr Ala Asn Ala Asp Pro
35 40 45
Glu Pro Ala Pro Pro Val Pro Thr Thr Ala Ala Ser Pro Pro Ser Thr
50 55 60
Ala Ala Ala Pro Pro Ala Pro Ala Thr Pro Val Ala Pro Pro Pro Pro
65 70 75 80
Ala Ala Ala Asn Thr Pro Asn Ala Gln Pro Gly Asp Pro Asn Ala Ala
85 90 95
Pro Pro Pro Ala Asp Pro Asn Ala Pro Pro Pro Pro Val Ile Ala Pro
100 105 110
Asn Ala pro Gln Pro Val Arg Ile Asp Asn Pro Val Gly Gly Phe Ser
115 120 125
Phe Ala Leu Pro Ala Gly Trp Val Glu Ser Asp Ala Ala His Phe Asp
130 135 140
Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr Gly Asp Pro Pro Phe Pro
145 150 155 160
Gly Gln Pro Pro Pro Val Ala Asn Asp Thr Arg Ile Val Leu Gly Arg
165 170 175
Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu Ala Thr Asp Ser Lys Ala
180 185 190
Ala Ala Arg Leu Gly Ser Asp Met Gly Glu Phe Tyr Met Pro Tyr Pro
195 200 205
Gly Thr Arg Ile Asn Gln Glu Thr Val Ser Leu Asp Ala Asn Gly Val
210 215 220
Ser Gly Ser Ala Ser Tyr Tyr Gln Val Lys Phe Ser Asp Pro Ser Lys
225 230 235 240
Pro Asn Gly Gln Ile Trp Thr Gly Val Ile Gly Ser Pro Ala Ala Asn
245 250 255
Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp Phe Val Val Trp Leu Gly
260 265 270
Thr Ala Asn Asn Pro Val Asp Lys Gly Ala Ala Lys Ala Leu Ala Glu
275 280 285
Ser Ile Arg Pro Leu Val Ala Pro Pro Pro Ala Pro Ala Pro Ala Pro
290 295 300
Ala Ala Pro Ala Pro Val Ala Pro Ala Gly Glu Val Ala Pro Thr
305 310 315
Pro Thr Thr Pro Thr Thr Ala Arg Thr Leu Pro Ala
320 325 330

INFORMATION FOR SEQ ID NO:54

SEQUENCE CHARACTERISTICS

- A) LENGTH, 20 AMINO ACIDS
- B) TYPE AMINO ACIDS
- C) STRANDEDNESS
- D) TOPOLOGY

1 5 10 15
Val Ala Ala Leu
20

(2) INFORMATION FOR SEQ ID NO:55:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Ala Val Glu Ser Gly Met Leu Ala Leu Gly Thr Pro Ala Pro Ser
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:56:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 19 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

Ala Ala Met Lys Pro Arg Thr Gly Asp Gly Pro Leu Glu Ala Ala Lys
 1 5 10 15
 Glu Gly Arg

(2) INFORMATION FOR SEQ ID NO:57:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:58:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 14 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:

Asp Ile Gly Ser Glu Ser Thr Glu Asp Gln Gln Xaa Ala Val
 1 5 10

(2) INFORMATION FOR SEQ ID NO:59:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 13 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

Ala Glu Glu Ser Ile Ser Thr Xaa Glu Xaa Ile Val Pro
 1 5 10

(2) INFORMATION FOR SEQ ID NO:60:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 17 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Asp Pro Glu Pro Ala Pro Pro Val Pro Thr Ala Ala Ala Ala Pro Pro
 1 5 10 15
 Ala

(2) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

Ala Pro Lys Thr Thr Xaa His His Lys Lys Gly Thr Asp Thr Gly
 1 5 10

(2) INFORMATION FOR SEQ ID NO:62:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

Leu Leu Asn Asn Leu Ala Asp Pro Asp Val Ser Phe Ala Asp
20 25 30

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 187 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

Thr	Gly	Ser	Leu	Asn	Gln	Thr	His	Asn	Arg	Arg	Ala	Asn	Glu	Arg	Lys
1				5										15	
Asn	Thr	Thr	Met	Lys	Met	Val	Lys	Ser	Ile	Ala	Ala	Gly	Leu	Thr	Ala
			20					25					30		
Ala	Ala	Ala	Ile	Gly	Ala	Ala	Ala	Ala	Gly	Val	Thr	Ser	Ile	Met	Ala
		35				40						45			
Gly	Gly	Pro	Val	Val	Tyr	Gln	Met	Gln	Pro	Val	Val	Phe	Gly	Ala	Pro
	50					55					60				
Leu	Pro	Leu	Asp	Pro	Ala	Ser	Ala	Pro	Asp	Val	Pro	Thr	Ala	Ala	Gln
65				70							75				
Leu	Thr	Ser	Leu	Leu	Asn	Ser	Leu	Ala	Asp	Pro	Asn	Val	Ser	Phe	Ala
			85						90					95	
Asn	Lys	Gly	Ser	Leu	Val	Glu	Gly	Gly	Ile	Gly	Gly	Thr	Glu	Ala	Arg
			100					105					110		
Ile	Ala	Asp	His	Lys	Leu	Lys	Lys	Ala	Ala	Glu	His	Gly	Asp	Leu	Pro
		115				120						125			
Leu	Ser	Phe	Ser	Val	Thr	Asn	Ile	Gln	Pro	Ala	Ala	Ala	Gly	Ser	Ala
		130				135						140			
Thr	Ala	Asp	Val	Ser	Val	Ser	Gly	Pro	Lys	Leu	Ser	Ser	Pro	Val	Thr
145				150						155					160
Gln	Asn	Val	Thr	Phe	Val	Asn	Gln	Gly	Gly	Trp	Met	Leu	Ser	Arg	Ala
			165						170						175
Ser	Ala	Met	Glu	Leu	Leu	Gln	Ala	Ala	Gly	Xaa					
			180					185							

2 INFORMATION FOR SEQ ID NO:64

SEQUENCE CHARACTERISTICS

- A LENGTH 148 amino acids
B TYPE: amino acid
C STRANDEDNESS single
D TOPOLOGY linear

X1. SEQUENCE DESCRIPTION: SEQ ID NO. 64

[illegible]

```

      50              55              60
Ala Gly Arg His Pro Asp Ser Asp Ile Phe Leu Asp Asp Val Thr Val
65              70              75              80
Ser Arg Arg His Ala Glu Phe Arg Leu Glu Asn Asn Glu Phe Asn Val
      85              90              95
Val Asp Val Gly Ser Leu Asn Gly Thr Tyr Val Asn Arg Glu Pro Val
      100             105             110
Asp Ser Ala Val Leu Ala Asn Gly Asp Glu Val Gln Ile Gly Lys Leu
      115             120             125
Arg Leu Val Phe Leu Thr Gly Pro Lys Gln Gly Glu Asp Asp Gly Ser
      130             135             140
Thr Gly Gly Pro
145

```

(2) INFORMATION FOR SEQ ID NO:65:

```

(1) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 230 amino acids
      (B) TYPE: amino acid
      (C) STRANDEDNESS: single
      (D) TOPOLOGY: linear

```

(1) SEQUENCE DESCRIPTION: SEQ ID NO:65:

```

Thr Ser Asn Arg Pro Ala Arg Arg Gly Arg Arg Ala Pro Arg Asp Thr
1      5      10      15
Gly Pro Asp Arg Ser Ala Ser Leu Ser Leu Val Arg His Arg Arg Gln
20      25      30
Gln Arg Asp Ala Leu Cys Leu Ser Thr Gln Ile Ser Arg Gln Ser
35      40      45
Asn Leu Pro Pro Ala Ala Gly Gly Ala Ala Asn Tyr Ser Arg Arg Asn
50      55      60
Phe Asp Val Arg Ile Lys Ile Phe Met Leu Val Thr Ala Val Val Leu
65      70      75      80
Leu Cys Cys Ser Gly Val Ala Thr Ala Ala Pro Lys Thr Tyr Cys Glu
85      90      95
Glu Leu Lys Gly Thr Asp Thr Gly Gln Ala Cys Gln Ile Gln Met Ser
100      105      110
Asp Pro Ala Tyr Asn Ile Asn Ile Ser Leu Pro Ser Tyr Tyr Pro Asp
115      120      125
Gln Lys Ser Leu Thr Asn Tyr Ile Ala Gln Thr Arg Asp Lys Phe Leu
130      135      140
Ser Ala Ala Thr Ser Ser Thr Pro Arg Glu Ala Pro Tyr Glu Leu Asn
145      150      155
Ile Thr Ser Ala Thr Tyr Gln Ser Ala Ile Pro Pro Arg Gly Thr Gln
160      165      170      175
Ala Val Val Leu Asn Val Tyr His Asn Ala Gly Gly Thr His Pro Thr
180      185      190
Thr Thr Thr Lys Ala Phe Asp Pro Asp Gln Ala Tyr Arg Lys Pro Ile
195      200      205

```

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 132 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

```

Thr Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gly Gln Gly Phe
1      5      10      15
Ala Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser
20      25      30
Gly Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly
35      40      45
Leu Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val
50      55      60
Val Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val
65      70      75      80
Ile Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala
85      90      95
Asp Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Asn Trp
100     105     110
Gln Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Glu
115     120     125
Gly Pro Pro Ala
130

```

2. INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 100 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

```

Val Pro Leu Arg Ser Pro Ser Met Ser Trp Ser Lys Tyr Leu Ala Ala
1      5      10      15
Ala His Arg Asn Pro Val Ile Arg Ala Arg Arg Leu Ser Asn Pro Pro
20      25      30
Pro Arg Lys Tyr Arg Ser Met Pro Ser Ser Ala Thr Ala Ser Ala Gly
35      40      45
Met Ala Arg Val Arg Arg Arg Ala Ile Trp Arg Gly Pro Ala Thr Xaa
50      55      60
Ser Ala Gly Met Ala Arg Val Arg Arg Trp Xaa Val Met Pro Xaa Val
65      70      75      80
Ile Gln Ser Thr Xaa Ile Arg His Xaa Gly Pro Ser Asp Asn Arg Gly
85      90      95
Val Ala Arg Lys
100

```


(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 163 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

```

Met Thr Asp Asp Ile Leu Leu Ile Asp Thr Asp Glu Arg Val Arg Thr
1           5           10           15
Leu Thr Leu Asn Arg Pro Gln Ser Arg Asn Ala Leu Ser Ala Ala Leu
20           25           30
Arg Asp Arg Phe Phe Ala Xaa Leu Xaa Asp Ala Glu Xaa Asp Asp Asp
35           40           45
Ile Asp Val Val Ile Leu Thr Gly Ala Asp Pro Val Phe Cys Ala Gly
50           55           60
Leu Asp Leu Lys Val Ala Gly Arg Ala Asp Arg Ala Ala Gly His Leu
65           70           75           80
Thr Ala Val Gly Gly His Asp Gln Ala Gly Asp Arg Arg Asp Gln Arg
85           90           95
Arg Arg Gly His Arg Arg Ala Arg Thr Gly Ala Val Leu Arg His Pro
100          105          110
Asp Arg Leu Arg Ala Arg Pro Leu Arg Arg His Pro Arg Pro Gly Gly
115          120          125
Ala Ala Ala His Leu Gly Thr Gln Cys Val Leu Ala Ala Lys Gly Arg
130          135          140
His Arg Xaa Gly Pro Val Asp Glu Pro Asp Arg Arg Leu Pro Val Arg
145          150          155          160
Asp Arg Arg

```

(i) INFORMATION FOR SEQ ID NO:69:

1. SEQUENCE CHARACTERISTICS

- (A) LENGTH: 344 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

xi. SEQUENCE DESCRIPTION: SEQ ID NO:69:

```

Met Lys Phe Val Asn His Ile Glu Pro Val Ala Pro Arg Arg Ala Gly
1           5           10           15
Ile Ala Val Ala Glu Val Tyr Ala Glu Ala Arg Arg Gln Phe Gly Arg
20           25           30
Leu Pro Glu Pro Leu Ala Met Leu Ser Pro Asp Glu Gly Leu Leu Thr
35           40           45
Ala Gly Tyr Ala Thr Leu Arg Glu Thr Leu Leu Val Gly Gln Val Pro
50           55           60
Arg Gly Arg Asp His Ala Val Ala Ala Ala Thr Ala Ala Ser Leu Arg
65           70           75           80

```

Gly Asp Pro Asn Ala Pro Tyr Val Ala Trp Ala Ala Gly Thr Gly Thr
 115 120 125
 Pro Ala Gly Pro Pro Ala Pro Phe Gly Pro Asp Val Ala Ala Glu Tyr
 130 135 140
 Leu Gly Thr Ala Val Gln Phe His Phe Ile Ala Arg Leu Val Leu Val
 145 150 155 160
 Leu Leu Asp Glu Thr Phe Leu Pro Gly Gly Pro Arg Ala Gln Gln Leu
 165 170 175
 Met Arg Arg Ala Gly Gly Leu Val Phe Ala Arg Lys Val Arg Ala Glu
 180 185 190
 His Arg Pro Gly Arg Ser Thr Arg Arg Leu Glu Pro Arg Thr Leu Pro
 195 200 205
 Asp Asp Leu Ala Trp Ala Thr Pro Ser Glu Pro Ile Ala Thr Ala Phe
 210 215 220
 Ala Ala Leu Ser His His Leu Asp Thr Ala Pro His Leu Pro Pro Pro
 225 230 235 240
 Thr Arg Gln Val Val Arg Arg Val Val Gly Ser Trp His Gly Glu Pro
 245 250 255
 Met Pro Met Ser Ser Arg Trp Thr Asn Glu His Thr Ala Glu Leu Pro
 260 265 270
 Ala Asp Leu His Ala Pro Thr Arg Leu Ala Leu Leu Thr Gly Leu Ala
 275 280 285
 Pro His Gln Val Thr Asp Asp Val Ala Ala Ala Arg Ser Leu Leu
 290 295 300
 Asp Thr Asp Ala Ala Leu Val Gly Ala Leu Ala Trp Ala Ala Phe Thr
 305 310 315 320
 Ala Ala Arg Arg Ile Gly Thr Trp Ile Gly Ala Ala Ala Glu Gly Gln
 325 330 335
 Val Ser Arg Gln Asn Pro Thr Gly
 340

01 INFORMATION FOR SEQ ID NO: 101:

1. SEQUENCE CHARACTERISTICS
 - (A) LENGTH: 485 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

02 SEQUENCE DESCRIPTION: SEQ ID NO: 101

Asp Asp Pro Asp Met Pro His Thr Val Ala Lys Ala Val Ala Asp Ala
 1 5 10 15
 Leu Gly Arg Gly Ile Ala Pro Val Gln Asp Ile Gln Asp Lys Val Gln
 20 25 30
 Ala Arg Leu Gly Gln Ala Gly Leu Asp Asp Val Ala Arg Val Tyr Ile
 35 40 45
 Ile Tyr Arg Gln Arg Arg Ala Gln Leu Arg Thr Ala Lys Ala Leu Leu
 50 55 60
 Gly Val Arg Asp Ala Leu Leu Thr Ser Leu Ala Ala Val Thr Val Leu
 65 70 75
 Val Val Val Val Val Val Val Val Val Val Val Val Val Val Val Val
 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485

Asp Gln Tyr Glu Pro Gly Ser Ser Arg Arg Trp Ala Glu Arg Phe Ala
 115 120 125
 Thr Leu Leu Arg Asn Leu Glu Phe Leu Pro Asn Ser Pro Thr Leu Met
 130 135 140
 Asn Ser Gly Thr Asp Leu Gly Leu Leu Ala Gly Cys Phe Val Leu Pro
 145 150 155 160
 Ile Glu Asp Ser Leu Gln Ser Ile Phe Ala Thr Leu Gly Gln Ala Ala
 165 170 175
 Glu Leu Gln Arg Ala Gly Gly Gly Thr Gly Tyr Ala Phe Ser His Leu
 180 185 190
 Arg Pro Ala Gly Asp Arg Val Ala Ser Thr Gly Gly Thr Ala Ser Gly
 195 200 205
 Pro Val Ser Phe Leu Arg Leu Tyr Asp Ser Ala Ala Gly Val Val Ser
 210 215 220
 Met Gly Gly Arg Arg Arg Gly Ala Cys Met Ala Val Leu Asp Val Ser
 225 230 235 240
 His Pro Asp Ile Cys Asp Phe Val Thr Ala Lys Ala Glu Ser Pro Ser
 245 250 255
 Glu Leu Pro His Phe Asn Leu Ser Val Gly Val Thr Asp Ala Phe Leu
 260 265 270
 Arg Ala Val Glu Arg Asn Gly Leu His Arg Leu Val Asn Pro Arg Thr
 275 280 285
 Gly Lys Ile Val Ala Arg Met Pro Ala Ala Glu Phe Asp Ala Ile
 290 295 300
 Cys Lys Ala Ala His Ala Gly Gly Asp Pro Gly Leu Val Phe Leu Asp
 305 310 315 320
 Thr Ile Asn Arg Ala Asn Pro Val Pro Gly Arg Gly Arg Ile Glu Ala
 325 330 335
 Thr Asn Pro Cys Gly Glu Val Pro Leu Leu Pro Tyr Glu Ser Cys Asn
 340 345 350
 Leu Gly Ser Ile Asn Leu Ala Arg Met Leu Ala Asp Gly Arg Val Asp
 355 360 365
 Trp Asp Arg Leu Gln His Val Ala Gly Val Ala Val Arg Phe Leu Asp
 370 375 380
 Asp Val Ile Asp Val Ser Arg Tyr Pro Phe Pro Glu Leu Gly Glu Ala
 385 390 395 400
 Ala Arg Ala Thr Arg Lys Ile Gly Leu Gly Val Met Gly Leu Ala Glu
 405 410 415
 Leu Leu Ala Ala Leu Gly Ile Pro Tyr Asp Ser Glu Glu Ala Val Arg
 420 425 430
 Leu Ala Thr Arg Leu Met Arg Arg Ile Gln Gln Ala Ala His Thr Ala
 435 440 445
 Ser Arg Arg Leu Ala His Thr Arg Tyr Ala Phe Pro Ala Phe Thr Asp
 450 455 460
 Ser Arg Phe Ala Arg Ser Gly Pro Arg Arg Asn Ala His Val Thr Ser
 465 470 475 480
 Val Ala Pro Thr Gly
 485

3 INFORMATION FOR SEQ 10 AND 11

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

Gly Val Ile Val Leu Asp Leu Glu Pro Arg Gly Pro Leu Pro Thr Glu
 1 5 10 15
 Ile Tyr Trp Arg Arg Arg Gly Leu Ala Leu Gly Ile Ala Val Val Val
 20 25 30
 Val Gly Ile Ala Val Ala Ile Val Ile Ala Phe Val Asp Ser Ser Ala
 35 40 45
 Gly Ala Lys Pro Val Ser Ala Asp Lys Pro Ala Ser Ala Gln Ser His
 50 55 60
 Pro Gly Ser Pro Ala Pro Gln Ala Pro Gln Pro Ala Gly Gln Thr Glu
 65 70 75 80
 Gly Asn Ala Ala Ala Pro Pro Gln Gly Gln Asn Pro Glu Thr Pro
 85 90 95
 Thr Pro Thr Ala Ala Val Gln Pro Pro Val Leu Lys Glu Gly Asp
 100 105 110
 Asp Cys Pro Asp Ser Thr Leu Ala Val Lys Gly Leu Thr Asn Ala Pro
 115 120 125
 Gln Tyr Tyr Val Gly Asp Gln Pro Lys Phe Thr Met Val Val Thr Asn
 130 135 140
 Ile Gly Leu Val Ser Cys Lys Arg Asp Val Gly Ala Ala Val Leu Ala
 145 150 155 160
 Ala Tyr Val Tyr Ser Leu Asp Asn Lys Arg Leu Trp Ser Asn Leu Asp
 165 170 175
 Cys Ala Pro Ser Asn Gln Thr Leu Val Lys Thr Phe Ser Pro Gly Glu
 180 185 190
 Gln Val Thr Thr Ala Val Thr Trp Thr Gly Met Gly Ser Ala Pro Arg
 195 200 205
 Cys Pro Leu Pro Arg Pro Ala Ile Gly Pro Gly Thr Tyr Asn Leu Val
 210 215 220
 Val Gln Leu Gly Asn Leu Ala Ser Leu Pro Val Pro Phe Ile Leu Asn
 225 230 235 240
 Gln Pro Pro Pro Pro Gly Pro Val Pro Ala Pro Gly Pro Ala Gln
 245 250 255
 Ala Pro Pro Pro Glu Ser Pro Ala Gln Gly Gly
 260 265

1 INFORMATION FOR SEQ ID NO:71:

- 1 SEQUENCE CHARACTERISTICS
 A LENGTH: 265 amino acids
 B TYPE: amino acid
 C STRANDEDNESS: single
 D TOPOLOGY: linear

xi SEQUENCE DESCRIPTION: SEQ ID NO:71:

Leu Ile Ser Thr Gly Asp Ala Ser His Ala Ser Leu Gly Val Gln Val
 1

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 364 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

Gly Ala Ala Val Ser Leu Leu Ala Ala Gly Thr Leu Val Leu Thr Ala
1 5 10 15
Cys Gly Gly Gly Thr Asn Ser Ser Ser Ser Gly Ala Gly Gly Thr Ser
20 25 30
Gly Ser Val His Cys Gly Gly Lys Lys Glu Leu His Ser Ser Gly Ser
35 40 45
Thr Ala Gln Glu Asn Ala Met Gln Gln Phe Val Tyr Ala Tyr Val Arg
50 55 60
Ser Cys Pro Gly Tyr Thr Leu Asp Tyr Asn Ala Asn Gly Ser Gly Ala
65 70 75 80
Gly Val Thr Gln Phe Leu Asn Asn Glu Thr Asp Phe Ala Gly Ser Asp
85 90 95
Val Pro Leu Asn Pro Ser Thr Gly Gln Pro Asp Arg Ser Ala Glu Arg
100 105 110
Lys Gly Ser Pro Ala Trp Asp Leu Pro Thr Val Phe Gly Pro Ile Ala.
115 120 125
Ile Thr Tyr Asn Ile Lys Gly Val Ser Thr Leu Asn Leu Asp Gly Pro
130 135 140
Thr Thr Ala Lys Ile Phe Asn Gly Thr Ile Thr Val Trp Asn Asp Pro
145 150 155 160
Gln Ile Gln Ala Leu Asn Ser Gly Thr Asp Leu Pro Pro Thr Pro Ile
165 170 175
Ser Val Ile Phe Arg Ser Asp Lys Ser Gly Thr Ser Asp Asn Phe Gln
180 185 190
Lys Tyr Leu Asp His Val Ser Asn Gly Ala Trp Gly Lys Gly Ala Ser
195 200 205
Glu Thr Phe Ser Gly Gly Val Gly Val Gly Ala Ser Gly Asn Asn Gly
210 215 220
Thr Ser Ala Leu Leu Gln Thr Thr Asp Gly Ser Ile Thr Tyr Asn Glu
225 230 235 240
Trp Ser Phe Ala Val Gly Lys Gln Leu Asn Met Ala Gln Ile Ile Thr
245 250 255

```

      290              295              300
Val Leu Ala Thr Tyr Glu Ile Val Cys Ser Lys Tyr Pro Asp Ala Thr
305              310              315              320
Thr Gly Thr Ala Val Arg Ala Phe Met Gln Ala Ala Ile Gly Pro Gly
      325              330              335
Gln Glu Gly Leu Asp Gln Tyr Gly Ser Ile Pro Leu Pro Lys Ser Phe
      340              345              350
Gln Ala Lys Leu Ala Ala Ala Val Asn Ala Ile Ser
      355              360

```

(2) INFORMATION FOR SEQ ID NO:74:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 309 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:74:

```

Gln Ala Ala Ala Gly Arg Ala Val Arg Arg Thr Gly His Ala Glu Asp
1              5              10              15
Gln Thr His Gln Asp Arg Leu His His Gly Cys Arg Arg Ala Ala Val
      20              25              30
Val Val Arg Gln Asp Arg Ala Ser Val Ser Ala Thr Ser Ala Arg Pro
      35              40              45
Pro Arg Arg His Pro Ala Gln Gly His Arg Arg Arg Val Ala Pro Ser
      50              55              60
Gly Gly Arg Arg Arg Pro His Pro His His Val Gln Pro Asp Asp Arg
      65              70              75              80
Arg Asp Arg Pro Ala Leu Leu Asp Arg Thr Gln Pro Ala Glu His Pro
      85              90              95
Asp Pro His Arg Arg Gly Pro Ala Asp Pro Gly Arg Val Arg Gly Arg
      100              105              110
Gly Arg Leu Arg Arg Val Asp Asp Gly Arg Leu Gln Pro Asp Arg Asp
      115              120              125
Ala Asp His Gly Ala Pro Val Arg Gly Arg Gly Pro His Arg Gly Val
      130              135              140
Gln His Arg Gly Gly Pro Val Phe Val Arg Arg Val Pro Gly Val Arg
      145              150              155              160
Cys Ala His Arg Arg Gly His Arg Arg Val Ala Ala Pro Gly Gln Gly
      165              170              175
Asp Val Leu Arg Ala Gly Leu Arg Val Gln Arg Leu Arg Pro Val Ala
      180              185              190
Ala Val Ala Asn Leu His Arg Gly Ser Gln Arg Ala Asp Gly Arg Val
      195              200              205
Phe Arg Pro Ile Arg Arg Gly Ala Arg Leu Pro Ala Arg Arg Ser Arg
      210              215              220
Ala Gly Pro Gln Gly Arg Leu His Leu Asn Gly Ala Gly Pro Ser Pro
      225              230              235              240
Leu Ser Ala Arg Ala Gly Gln Gln Gly Val Val Val Val Val Val Val

```


305 310 315 320
 Pro Pro Leu Gly Ser Pro Pro Pro Lys Thr Thr Val Thr Phe Ser Asp
 325 330 335
 Gly Arg Thr Ala Pro Phe Thr Val Val Gly Ala Asp Pro Thr Ser Asp
 340 345 350
 Ile Ala Val Val Arg Val Gln Gly Val Ser Gly Leu Thr Pro Ile Ser
 355 360 365
 Leu Gly Ser Ser Ser Asp Leu Arg Val Gly Gln Pro Val Leu Ala Ile
 370 375 380
 Gly Ser Pro Leu Gly Leu Glu Gly Thr Val Thr Thr Gly Ile Val Ser
 385 390 395 400
 Ala Leu Asn Arg Pro Val Ser Thr Thr Gly Glu Ala Gly Asn Gln Asn
 405 410 415
 Thr Val Leu Asp Ala Ile Gln Thr Asp Ala Ala Ile Asn Pro Gly Asn
 420 425 430
 Ser Gly Gly Ala Leu Val Asn Met Asn Ala Gln Leu Val Gly Val Asn
 435 440 445
 Ser Ala Ile Ala Thr Leu Gly Ala Asp Ser Ala Asp Ala Gln Ser Gly
 450 455 460
 Ser Ile Gly Leu Gly Phe Ala Ile Pro Val Asp Gln Ala Lys Arg Ile
 465 470 475 480
 Ala Asp Glu Leu Ile Ser Thr Gly Lys Ala Ser His Ala Ser Leu Gly
 485 490 495
 Val Gln Val Thr Asn Asp Lys Asp Thr Pro Gly Ala Lys Ile Val Glu
 500 505 510
 Val Val Ala Gly Gly Ala Ala Ala Asn Ala Gly Val Pro Lys Gly Val
 515 520 525
 Val Val Thr Lys Val Asp Asp Arg Pro Ile Asn Ser Ala Asp Ala Leu
 530 535 540
 Val Ala Ala Val Arg Ser Lys Ala Pro Gly Ala Thr Val Ala Leu Thr
 545 550 555 560
 Phe Gln Asp Pro Ser Gly Gly Ser Arg Thr Val Gln Val Thr Leu Gly
 565 570 575
 Lys Ala Glu Gln
 580

INFORMATION FOR SEQ ID NO:76

1. SEQUENCE CHARACTERISTICS:
 A. LENGTH: 580 amino acids
 B. TYPE: amino acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:76

Met Asn Asp Gly Lys Arg Ala Val Thr Ser Ala Val Leu Val Val Leu
 1 5 10 15
 Gly Ala Cys Leu Ala Leu Arg Leu Ser Gly Cys Ser Ser Pro Lys Pro
 20 25 30
 Asp Ala Ala Glu Gln Ala Val Pro Val Ser Pro Thr Ala Ser Arg Leu


```

65          70          75          80
Leu Gly Ile Thr Ser Ala Asp Val Asp Val Arg Ala Asn Pro Leu Ala
          85          90          95
Ala Lys Gly Val Cys Thr Tyr Asn Asp Glu Gln Gly Val Pro Phe Arg
          100          105          110
Val Gln Gly Asp Asn Ile Ser Val Lys Leu Phe Asp Asp Trp Ser Asn
          115          120          125
Leu Gly Ser Ile Ser Glu Leu Ser Thr Ser Arg Val Leu Asp Pro Ala
          130          135          140
Ala Gly Val Thr Gln Leu Leu Ser Gly Val Thr Asn Leu Gln Ala Gln
145          150          155          160
Gly Thr Glu Val Ile Asp Gly Ile Ser Thr Thr Lys Ile Thr Gly Thr
          165          170          175
Ile Pro Ala Ser Ser Val Lys Met Leu Asp Pro Gly Ala Lys Ser Ala
          180          185          190
Arg Pro Ala Thr Val Trp Ile Ala Gln Asp Gly Ser His His Leu Val
          195          200          205
Arg Ala Ser Ile Asp Leu Gly Ser Gly Ser Ile Gln Leu Thr Gln Ser
          210          215          220
Lys Trp Asn Glu Pro Val Asn Val Asp
225          230

```

(2) INFORMATION FOR SEQ ID NO:77:

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 66 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:77:

```

Val Ile Asp Ile Ile Gly Thr Ser Pro Thr Ser Trp Gln Gln Ala Ala
1      5      10      15
Ala Glu Ala Val Gln Arg Ala Arg Asp Ser Val Asp Asp Ile Arg Val
20      25      30
Ala Arg Val Ile Glu Gln Asp Met Ala Val Asp Ser Ala Gly Lys Ile
35      40      45
Thr Tyr Arg Ile Lys Leu Glu Val Ser Phe Lys Met Arg Pro Ala Gln
50      55      60
Pro Arg
66

```

(3) INFORMATION FOR SEQ ID NO:78:

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 69 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:78:

Cys Ala Ser Pro Pro Ser Pro Pro Leu Pro Pro Ala Pro Pro Val Ala
 20 25 30
 Pro Gly Pro Pro Met Pro Pro Leu Asp Pro Trp Pro Pro Ala Pro Pro
 35 40 45
 Leu Pro Tyr Ser Thr Pro Pro Gly Ala Pro Leu Pro Pro Ser Pro Pro
 50 55 60
 Ser Pro Pro Leu Pro
 65

(2) INFORMATION FOR SEQ ID NO.79:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 355 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) SEQUENCE DESCRIPTION: SEQ ID NO:79:

Met Ser Asn Ser Arg Arg Arg Ser Leu Arg Trp Ser Trp Leu Leu Ser
 1 5 10 15
 Val Leu Ala Ala Val Gly Leu Gly Leu Ala Thr Ala Pro Ala Gln Ala
 20 25 30
 Ala Pro Pro Ala Leu Ser Gln Asp Arg Phe Ala Asp Phe Pro Ala Leu
 35 40 45
 Pro Leu Asp Pro Ser Ala Met Val Ala Gln Val Ala Pro Gln Val Val
 50 55 60
 Asn Ile Asn Thr Lys Leu Gly Tyr Asn Asn Ala Val Gly Ala Gly Thr
 65 70 75 80
 Gly Ile Val Ile Asp Pro Asn Gly Val Val Leu Thr Asn Asn His Val
 85 90 95
 Ile Ala Gly Ala Thr Asp Ile Asn Ala Phe Ser Val Gly Ser Gly Gln
 100 105 110
 Thr Tyr Gly Val Asp Val Val Gly Tyr Asp Arg Thr Gln Asp Val Ala
 115 120 125
 Val Leu Gln Leu Arg Gly Ala Gly Gly Leu Pro Ser Ala Ala Ile Gly
 130 135 140
 Gly Gly Val Ala Val Gly Gly Pro Val Val Ala Met Gly Asn Ser Gly
 145 150 155 160
 Gly Gln Gly Gly Thr Pro Arg Ala Val Pro Gly Arg Val Val Ala Leu
 165 170 175
 Gly Gln Thr Val Gln Ala Ser Asp Ser Leu Thr Gly Ala Gln Gln Thr
 180 185 190
 Leu Asn Gly Leu Ile Gln Phe Asp Ala Ala Ile Gln Pro Gly Asp Ser
 195 200 205
 Gly Gly Pro Val Val Asn Gly Leu Gly Gln Val Val Gly Met Asn Thr
 210 215 220
 Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gly Gln Gly Phe Ala
 225 230 235 240
 Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser Gly
 245 250 255
 Leu Ile Ser Asn Val Val

Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val Ile
 290 295 300
 Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala Asp
 305 310 315 320
 Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Asn Trp Gln
 325 330 335
 Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Glu Gly
 340 345 350
 Pro Pro Ala
 355

(2) INFORMATION FOR SEQ ID NO:80:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 205 amino acids

(B) TYPE: amino acid

(C) STANNEDNESS: single

(D) TOPOLOGY: linear

(11) SEQUENCE DESCRIPTION: SEQ ID NO:80:

Ser Pro Lys Pro Asp Ala Glu Glu Gln Gly Val Pro Val Ser Pro Thr
 1 5 10 15
 Ala Ser Asp Pro Ala Leu Leu Ala Glu Ile Arg Gln Ser Leu Asp Ala
 20 25 30
 Thr Lys Gly Leu Thr Ser Val His Val Ala Val Arg Thr Thr Gly Lys
 35 40 45
 Val Asp Ser Leu Leu Gly Ile Thr Ser Ala Asp Val Asp Val Arg Ala
 50 55 60
 Asn Pro Leu Ala Ala Lys Gly Val Cys Thr Tyr Asn Asp Glu Gln Gly
 65 70 75 80
 Val Pro Phe Arg Val Gln Gly Asp Asn Ile Ser Val Lys Leu Phe Asp
 85 90 95
 Asp Trp Ser Asn Leu Gly Ser Ile Ser Glu Leu Ser Thr Ser Arg Val
 100 105 110
 Leu Asp Pro Ala Ala Gly Val Thr Gln Leu Leu Ser Gly Val Thr Asn
 115 120 125
 Leu Gln Ala Gln Gly Thr Glu Val Ile Asp Gly Ile Ser Thr Thr Lys
 130 135 140
 Ile Thr Gly Thr Ile Pro Ala Ser Ser Val Lys Met Leu Asp Pro Gly
 145 150 155 160
 Ala Lys Ser Ala Arg Pro Ala Thr Val Trp Ile Ala Gln Asp Gly Ser
 165 170 175
 His His Leu Val Arg Ala Ser Ile Asp Leu Gly Ser Gly Ser Ile Gln
 180 185 190
 Leu Thr Gln Ser Lys Trp Asn Gln Pro Val Asn Val Asp
 195 200 205

(2) INFORMATION FOR SEQ ID NO:81:

(1) SEQUENCE CHARACTERISTICS:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:81:

Gly Asp Ser Phe Trp Ala Ala Ala Asp Gln Met Ala Arg Gly Phe Val
 1 5 10 15
 Leu Gly Ala Thr Ala Gly Arg Thr Thr Leu Thr Gly Glu Gly Leu Gln
 20 25 30
 His Ala Asp Gly His Ser Leu Leu Leu Asp Ala Thr Asn Pro Ala Val
 35 40 45
 Val Ala Tyr Asp Pro Ala Phe Ala Tyr Glu Ile Gly Tyr Ile Xaa Glu
 50 55 60
 Ser Gly Leu Ala Arg Met Cys Gly Glu Asn Pro Glu Asn Ile Phe Phe
 65 70 75 80
 Tyr Ile Thr Val Tyr Asn Glu Pro Tyr Val Gln Pro Pro Glu Pro Glu
 85 90 95
 Asn Phe Asp Pro Glu Gly Val Leu Gly Gly Ile Tyr Arg Tyr His Ala
 100 105 110
 Ala Thr Gln Gln Arg Thr Asn Lys Xaa Gln Ile Leu Ala Ser Gly Val
 115 120 125
 Ala Met Pro Ala Ala Leu Arg Ala Ala Gln Met Leu Ala Ala Glu Trp
 130 135 140
 Asp Val Ala Ala Asp Val Trp Ser Val Thr Ser Trp Gly Glu Leu Asn
 145 150 155 160
 Arg Asp Gly Val Val Ile Glu Thr Glu Lys Leu Arg His Pro Asp Arg
 165 170 175
 Pro Ala Gly Val Pro Tyr Val Thr Arg Ala Leu Glu Asn Ala Arg Gly
 180 185 190
 Pro Val Ile Ala Val Ser Asp Trp Met Arg Ala Val Pro Glu Gln Ile
 195 200 205
 Arg Pro Trp Val Pro Gly Thr Tyr Leu Thr Leu Gly Thr Asp Gly Phe
 210 215 220
 Gly Phe Ser Asp Thr Arg Pro Ala Gly Arg Arg Tyr Phe Asn Thr Asp
 225 230 235 240
 Ala Gln Ser Gln Val Gly Arg Gly Phe Gly Arg Gly Trp Pro Gly Arg
 245 250 255
 Arg Val Asn Ile Asp Pro Phe Gly Ala Gly Arg Gly Pro Pro Ala Gln
 260 265 270
 Leu Pro Gly Phe Asp Gln Gly Gly Gly Leu Arg Pro Xaa Lys
 275 280 285

INFORMATION FOR SEQ ID NO:81:

1. SEQUENCE CHARACTERISTICS:
 A. LENGTH: 285 amino acids
 B. TYPE: amino acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

X1 SEQUENCE DESCRIPTION: SEQ ID NO:81:

Thr Tyr Phe His Ala Leu Met Gln Thr Gln Ile His Asn Glu Ser Ser

111

```

      35              40              45
Asn His Ala Met Met Leu Val Gln His Leu Leu Asp Arg Asp Leu Arg
 50              55              60
Val Glu Ile Pro Gly Val Asp Thr Val Arg Asn Gln Phe Asp Arg Pro
 65              70              75              80
Arg Glu Ala Leu Ala Leu Ala Leu Asp Gln Glu Arg Thr Val Thr Asp
      85              90              95
Gln Val Gly Arg Leu Thr Ala Val Ala Arg Asp Glu Gly Asp Phe Leu
 100              105              110
Gly Glu Gln Phe Met Gln Trp Phe Leu Gln Glu Gln Ile Glu Glu Val
 115              120              125
Ala Leu Met Ala Thr Leu Val Arg Val Ala Asp Arg Ala Gly Ala Asn
 130              135              140
Leu Phe Glu Leu Glu Asn Phe Val Ala Arg Glu Val Asp Val Ala Pro
 145              150              155              160
Ala Ala Ser Gly Ala Pro His Ala Ala Gly Gly Arg Leu
      165              170

```

3. INFORMATION FOR SEQ ID NO:83:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 107 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:83:

```

Arg Ala Asp Glu Arg Lys Asn Thr Thr Met Lys Met Val Lys Ser Ile
 1              5              10              15
Ala Ala Gly Leu Thr Ala Ala Ala Ala Ile Gly Ala Ala Ala Gly
 20              25              30
Val Thr Ser Ile Met Ala Gly Gly Pro Val Val Tyr Gln Met Gln Pro
 35              40              45
Val Val Phe Gly Ala Pro Leu Pro Leu Asp Pro Xaa Ser Ala Pro Xaa
 50              55              60
Val Pro Thr Ala Ala Gln Trp Thr Xaa Leu Leu Asn Xaa Leu Xaa Asp
 65              70              75              80
Pro Asn Val Ser Phe Xaa Asn Lys Gly Ser Leu Val Gln Gly Gly Ile
 85              90              95
His Gly Xaa Gln Gly Xaa Xaa Arg Arg Xaa Gln
 100              105

```

4. INFORMATION FOR SEQ ID NO:84:

1. SEQUENCE CHARACTERISTICS:
- A. LENGTH: 126 amino acids
 - B. TYPE: amino acid
 - C. STRANDEDNESS: single
 - D. TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:84:

```

      20      25      30
Arg Arg Ala Leu Glu Leu Gln Ala Pro Ser Val Val Xaa Arg Gln Gly
  35      40      45
Val Lys Glu Pro Leu Xaa Thr Gly Ile Lys Ala Ile Asp Ala Met Thr
  50      55      60
Pro Ile Gly Arg Gly Gln Arg Gln Leu Ile Ile Gly Asp Arg Lys Thr
  65      70      75      80
Gly Lys Asn Arg Arg Leu Cys Arg Thr Pro Ser Ser Asn Gln Arg Glu
      85      90      95
Glu Leu Gly Val Arg Trp Ile Pro Arg Ser Arg Cys Ala Cys Val Tyr
      100      105      110
Val Gly His Arg Ala Arg Arg Gly Thr Tyr His Arg Arg
      115      120      125

```

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 117 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

```

Cys Asp Ala Val Met Gly Phe Leu Gly Gly Ala Gly Pro Leu Ala Val
  1      5      10      15
Val Asp Gln Gln Leu Val Thr Arg Val Pro Gln Gly Trp Ser Phe Ala
      20      25      30
Gln Ala Ala Ala Val Pro Val Val Phe Leu Thr Ala Trp Tyr Gly Leu
      35      40      45
Ala Asp Leu Ala Glu Ile Lys Ala Gly Glu Ser Val Leu Ile His Ala
      50      55      60
Gly Thr Gly Gly Val Gly Met Ala Ala Val Gln Leu Ala Arg Gln Trp
      65      70      75      80
Gly Val Glu Val Phe Val Thr Ala Ser Arg Gly Lys Trp Asp Thr Leu
      85      90      95
Arg Ala Xaa Xaa Phe Asp Asp Xaa Pro Tyr Arg Xaa Phe Pro His Xaa
      100      105      110
Arg Ser Ser Xaa Gly
      115

```

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 108 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

Met Thr Arg Pro ...

Thr Ala Pro Val Pro Asp Tyr Tyr Trp Cys Pro Gly Gln Pro Phe Asp
 35 40 45
 Pro Ala Trp Gly Pro Asn Trp Asp Pro Tyr Thr Cys His Asp Asp Phe
 50 55 60
 His Arg Asp Ser Asp Gly Pro Asp His Ser Arg Asp Tyr Pro Gly Pro
 65 70 75 80
 Ile Leu Glu Gly Pro Val Leu Asp Asp Pro Gly Ala Ala Pro Pro Pro
 85 90 95
 Pro Ala Ala Gly Gly Ala
 100

(2) INFORMATION FOR SEQ ID NO:87:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 88 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) SEQUENCE DESCRIPTION: SEQ ID NO:87:

Val Gln Cys Arg Val Trp Leu Glu Ile Gln Trp Arg Gly Met Leu Gly
 1 5 10 15
 Ala Asp Gln Ala Arg Ala Gly Gly Pro Ala Arg Ile Trp Arg Glu His
 20 25 30
 Ser Met Ala Ala Met Lys Pro Arg Thr Gly Asp Gly Pro Leu Glu Ala
 35 40 45
 Thr Lys Glu Gly Arg Gly Ile Val Met Arg Val Pro Leu Glu Gly Gly
 50 55 60
 Gly Arg Leu Val Val Glu Leu Thr Pro Asp Glu Ala Ala Ala Leu Gly
 65 70 75 80
 Asp Glu Leu Lys Gly Val Thr Ser
 85

(2) INFORMATION FOR SEQ ID NO:88:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 95 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) SEQUENCE DESCRIPTION: SEQ ID NO:88:

Thr Asp Ala Ala Thr Leu Ala Gln Gln Ala Gly Asn Phe Glu Arg Ile
 1 5 10 15
 Ser Gly Asp Leu Lys Thr Gln Ile Asp Gln Val Glu Ser Thr Ala Gly
 20 25 30
 Ser Leu Gln Gly Gln Thr Arg Gly Ala Ala Gly Thr Ala Ala Gln Ala
 35 40 45
 Ala Val Val Arg Phe Pro Gln Ala Ala Ala Ala Ala Ala Ala Ala
 50 55 60 65 70 75 80 85 90 95

85

90

95

(2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 166 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

```

Met Thr Gln Ser Gln Thr Val Thr Val Asp Gln Gln Glu Ile Leu Asn
1           5           10           15
Arg Ala Asn Glu Val Glu Ala Pro Met Ala Asp Pro Pro Thr Asp Val
20           25           30
Pro Ile Thr Pro Cys Glu Leu Thr Xaa Xaa Lys Asn Ala Ala Gln Gln
35           40           45
Xaa Val Leu Ser Ala Asp Asn Met Arg Glu Tyr Leu Ala Ala Gly Ala
50           55           60
Lys Glu Arg Gln Arg Leu Ala Thr Ser Leu Arg Asn Ala Ala Lys Xaa
65           70           75           80
Tyr Gly Glu Val Asp Glu Glu Ala Ala Thr Ala Leu Asp Asn Asp Gly
85           90           95
Glu Gly Thr Val Gln Ala Glu Ser Ala Gly Ala Val Gly Gly Asp Ser
100          105          110
Ser Ala Glu Leu Thr Asp Thr Pro Arg Val Ala Thr Ala Gly Glu Pro
115          120          125
Asn Phe Met Asp Leu Lys Glu Ala Ala Arg Lys Leu Glu Thr Gly Asp
130          135          140
Gln Gly Ala Ser Leu Ala His Xaa Gly Asp Gly Tyr Asn Thr Xaa Thr
145          150          155          160
Leu Thr Leu Gln Gly Asp
165

```

(2) INFORMATION FOR SEQ ID NO:90:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

```

Arg Ala Glu Arg Met
1           5

```

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

Val Ala Trp Met Ser Val Thr Ala Gly Gln Ala Glu Leu Thr Ala Ala
 1 5 10 15
 Gln Val Arg Val Ala Ala Ala Tyr Glu Thr Ala Tyr Gly Leu Thr
 20 25 30
 Val Pro Pro Pro Val Ile Ala Glu Asn Arg Ala Glu Leu Met Ile Leu
 35 40 45
 Ile Ala Thr Asn Leu Leu Gly Gln Asn Thr Pro Ala Ile Ala Val Asn
 50 55 60
 Glu Ala Glu Tyr Gly Glu Met Trp Ala Gln Asp Ala Ala Ala Met Phe
 65 70 75 80
 Gly Tyr Ala Ala Ala Thr Ala Thr Ala Thr Ala Thr Leu Leu Pro Phe
 85 90 95
 Glu Glu Ala Pro Glu Met Thr Ser Ala Gly Gly Leu Leu Gln Gln Ala
 100 105 110
 Ala Ala Val Glu Glu Ala Ser Asp Thr Ala Ala Ala Asn Gln Leu Met
 115 120 125
 Asn Asn Val Pro Gln Ala Leu Lys Gln Leu Ala Gln Pro Thr Gln Gly
 130 135 140
 Thr Thr Pro Ser Ser Lys Leu Gly Gly Leu Trp Lys Thr Val Ser Pro
 145 150 155 160
 His Arg Ser Pro Ile Ser Asn Met Val Ser Met Ala Asn Asn His Met
 165 170 175
 Ser Met Thr Asn Ser Gly Val Ser Met Thr Asn Thr Leu Ser Ser Met
 180 185 190
 Leu Lys Gly Phe Ala Pro Ala Ala Ala Gln Ala Val Gln Thr Ala
 195 200 205
 Ala Gln Asn Gly Val Arg Ala Met Ser Ser Leu Gly Ser Ser Leu Gly
 210 215 220
 Ser Ser Gly Leu Gly Gly Gly Val Ala Ala Asn Leu Gly Arg Ala Ala
 225 230 235 240
 Ser Val Arg Tyr Gly His Arg Asp Gly Gly Lys Tyr Ala Xaa Ser Gly
 245 250 255
 Arg Arg Asn Gly Gly Pro Ala
 260

2 INFORMATION FOR SEQ ID NO:92:

1. SEQUENCE CHARACTERISTICS
 A. LENGTH: 263 amino acids
 B. TYPE: amino acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

xi SEQUENCE DESCRIPTION: SEQ ID NO:92:

Met Thr Tyr Ser Pro Ala Asn Ile Gly Tyr Phe Gln Ala Gln Pro Ala
 1 10 15
 Gly Ser Thr ...
 20 25

50 55 60
 Glu Leu Gly Gly Gly Asp Gly Ala Val Ser Gly Asp Thr Gly Leu Pro
 65 70 75 80
 Val Gly Val Ala Leu Leu Ala Ala Leu Leu Ala Gly Val Val Leu Val
 85 90 95
 Pro Lys Ala Lys Ser His Val Thr Val Val Ala Val Leu Gly Val Leu
 100 105 110
 Gly Val Phe Leu Met Val Ser Ala Thr Phe Asn Lys Pro Ser Ala Tyr
 115 120 125
 Ser Thr Gly Trp Ala Leu Trp Val Val Leu Ala Phe Ile Val Phe Gln
 130 135 140
 Ala Val Ala Ala Val Leu Ala Leu Leu Val Glu Thr Gly Ala Ile Thr
 145 150 155 160
 Ala Pro Ala Pro Arg Pro Lys Phe Asp Pro Tyr Gly Gln Tyr Gly Arg
 165 170 175
 Tyr Gly Gln Tyr Gly Gln Tyr Gly Val Gln Pro Gly Gly Tyr Tyr Gly
 180 185 190
 Gln Gln Gly Ala Gln Gln Ala Ala Gly Leu Gln Ser Pro Gly Pro Gln
 195 200 205
 Gln Ser Pro Gln Pro Pro Gly Tyr Gly Ser Gln Tyr Gly Gly Tyr Ser
 210 215 220
 Ser Ser Pro Ser Gln Ser Gly Ser Gly Tyr Thr Ala Gln Pro Pro Ala
 225 230 235 240
 Gln Pro Pro Ala Gln Ser Gly Ser Gln Gln Ser His Gln Gly Pro Ser
 245 250 255
 Thr Pro Pro Thr Gly Phe Pro Ser Phe Ser Pro Pro Pro Val Ser
 260 265 270
 Ala Gly Thr Gly Ser Gln Ala Gly Ser Ala Pro Val Asn Tyr Ser Asn
 275 280 285
 Pro Ser Gly Gly Glu Gln Ser Ser Ser Pro Gly Gly Ala Pro Val
 290 295 300

2. INFORMATION FOR SEQ ID NO:93:

- (A) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 39 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

3. SEQUENCE DESCRIPTION (SEQ ID NO:93)

Gly Cys Gly Gln Thr Asp Ala Ala Thr Leu Ala Gln Glu Ala Gly Asn
 1 5 10 15
 Phe Glu Ala Thr Ser Gly Asp Leu Lys Thr Gln Ile
 20 25

4. INFORMATION FOR SEQ ID NO:94

- (A) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:94:

Asp Gln Val Glu Ser Thr Ala Gly Ser Leu Gln Gly Gln Trp Arg Gly
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:95:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:95:

Gly Cys Gly Ser Thr Ala Gly Ser Leu Gln Gly Gln Trp Arg Gly Ala
1 5 10 15
Ala Gly Thr Ala Ala Gln Ala Ala Val Val Arg
20 25

(2) INFORMATION FOR SEQ ID NO:96:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:96:

Gly Cys Gly Gly Thr Ala Ala Gln Ala Ala Val Val Arg Phe Gln Glu
1 5 10 15
Ala Ala Asn Lys Gln Lys Gln Glu Leu Asp Gln
20 25

(2) INFORMATION FOR SEQ ID NO:97:

(1) SEQUENCE CHARACTERISTICS:

- A: LENGTH: 27 amino acids
- B: TYPE: amino acid
- C: STRANDEDNESS: single
- D: TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:97:

Gly Cys Gly Ala Asn Lys Gln Lys Gln Glu Leu Asp Gln Ile Ser Thr
1 5 10 15
Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg
20 25

- (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(XI) SEQUENCE DESCRIPTION: SEQ ID NO:98:

Gly Cys Gly Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg Ala Asp Glu
 1 5 10 15
 Glu Gln Gln Gln Ala Leu Ser Ser Gln Met Gly Phe
 20 25

(2) INFORMATION FOR SEQ ID NO:99:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 507 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(XI) SEQUENCE DESCRIPTION: SEQ ID NO:99:

ATGAAGATGG TGAAATCGAT CGCCGCAGGT CTGACCGCCG CGGCTGCAAT CGGCGCCGCT 60
 GCGGCCCGGTG TGACTTCGAT CATGGCTGGG GCGCCGGTGG TATACCAGAT GCAGCCGGTC 120
 GTCTTGGGCG CGCCACTGCC GTTGGACCGG GCATCCGCGC CTGACGTCCC GACCGCCGCC 180
 CAGTTGACCA GCTTGCTCAA CAGCCTCCCG GATCCCAACG TGTCGTTTGC GAACAAGGGC 240
 AGTCTGTGTC AGGGCGGGCAT CGGGGGCACC GAGGCGCGCA TCGCCGACCA CAAGCTGAAG 300
 AAGGCGCGCG AGCACGGGGA TCTGCCGCTG TCGTTCAGCG TGACGAACAT CCAGCCGGCG 360
 GCGGCGGGTT CGGCCACCGC CGACGTTTCC GTCTGGGGTC CGAAGCTCTC GTCGCCGGTC 420
 ACGCAGAACG TCACGTCGT GAATCAAGGC AGCTGGATGC TGTACCGCGC ATCGCGGATG 480
 GAGTTGCTGC AGGCGCGAGG GAACTGA 507

(2) INFORMATION FOR SEQ ID NO:100:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 158 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(XI) SEQUENCE DESCRIPTION: SEQ ID NO:100:

Met Lys Met Val Lys Ser Ile Ala Ala Gly Ser Thr Ala Ala Ala Ala
 1 5 10 15
 Ile Gly Ala Ala Ala Ala Gly Val Thr Ser Ile Met Ala Gly Gly Pro
 20 25 30
 Val Val Tyr Glu Met Glu Pro Val Val Pro Gly Ala Pro Leu Pro Leu
 35 40 45
 Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Glu Ser Thr Ser
 50 55 60
 Leu Leu Asp Ser Leu Ala Asp Pro Val Thr Thr Ser Thr Thr Thr Thr

100 105 110
 Ser Val Thr Asn Ile Gln Pro Ala Ala Ala Gly Ser Ala Thr Ala Asp
 115 120 125
 Val Ser Val Ser Gly Pro Lys Leu Ser Ser Pro Val Thr Gln Asn Val
 130 135 140
 Thr Phe Val Asn Gln Gly Gly Trp Met Leu Ser Arg Ala Ser Ala Met
 145 150 155 160
 Glu Leu Leu Gln Ala Ala Gly Asn
 165

(2) INFORMATION FOR SEQ ID NO:101:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 500 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

X17 SEQUENCE DESCRIPTION: SEQ ID NO:101:

CGTGGCAATG TCGTTGACCG TCGGGGCCCG GGTGGCCTCC GCAGATCCCG TGGACGCGGT 60
 CATTAACACC ACCTGCAATT ACGGGCAGGT AGTAGCTGCG CTCAACGCGA CGGATCCGGG 120
 GGCTGCGCGA CAGTTCAACS CCTCACCAGT GCCCGAGTCC TATTGCGCA ATTTCTTCGC 180
 CGCACCAGCA CCTCAGCGCG CTGCCATGGC CGCGCAATTG CAAGCTGTGC CGGGGGCGGC 240
 ACACTACATC GGCCTTGTCG AGTCGGTTGC CGGCTCCTGC AACAACTATT AAGCCCCATGC 300
 GGGCCCCATC CCGCGACCCG GCATCGTCGC CGGGGCTAGG CCAGATTGCC CCGCTCCTCA 360
 ACGGGCGCGA TCCCGCGACC CCGCATCGTC CGCGGGGCTA GGCCAGATTG CCGCGCTCCT 420
 GAACGGGGCG CATCTCGTGC CGAATTCCTC CAGCCCCGGG GATCCACTAG TTCTAGAGCG 480
 GCGCGCACCG CGGTGAGCT

(3) INFORMATION FOR SEQ ID NO:102:

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 96 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

X18 SEQUENCE DESCRIPTION: SEQ ID NO:102:

Val Ala Met Ser Leu Thr Val Gly Ala Gly Val Ala Ser Ala Asn Pro
 1 5 10 15
 Val Asp Ala Val Ala Asn Thr Thr Lys Asn Tyr Gly Gln Val Val Ala
 20 25 30
 Ala Leu Asn Ala Thr Asp Pro Gly Ala Ala Ala Gln Phe Asn Ala Ser
 35 40 45
 Pro Val Ala Gln Ser Tyr Leu Arg Asn Phe Leu Ala Ala Pro Pro Pro
 50 55 60
 Gln Arg Ala Ala Met Ala Ala Gln Leu Gln Ala Val Pro Gly Ala Ala
 65 70 75 80 85 90 96

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 154 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:103:

```

ATGACAGAGC AGCACTGGAA TTTCGCGGGT ATCGAGGCCC CGGCAAGCGC AATCCAGGGA      60
AATGTCACGT CCATTCAATC CCTCCTTGAC GAGGGGAAGC AGTCCCTGAC CAAGCTCGCA      120
GCGGCCTGGG GCGGTAGCGG TTCGGAAGCG TACC                                154

```

(2) INFORMATION FOR SEQ ID NO:104:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 51 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:104:

```

Met Thr Glu Gln Gln Trp Asn Phe Ala Gly Ile Glu Ala Ala Ala Ser
 1             5             10             15
Ala Ile Gln Gly Asn Val Thr Ser Ile His Ser Leu Leu Asp Glu Gly
      20             25             30
Lys Gln Ser Leu Thr Lys Leu Ala Ala Ala Trp Gly Gly Ser Gly Ser
      35             40             45
Glu Ala Tyr
      50

```

(2) INFORMATION FOR SEQ ID NO:105:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 292 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:105:

```

TCTTCTGCGCA TTTCAGGCTT ACTATGAAGT TCGGCTTCAI TCTTGAAGAT TCTGAAAGCT      60
TCAATCTGAGT TCAATAACTG AGGTGCAATC TTAGCGGACT TTTCGAGAAG ATCCTGACCG      120
TCTGAAAGAG CGGACAGCGG GAGGCTGCTT TCGGCTAGG TCTGCTGCA AATCCCTCA      180
TCAATCTGAG TCGGGGCGCG TACAAGGAAG TCGGCTGCTA ATTGCTGCTG TATCTGCTCG      240
ACTTCTGCTG TCTGAGCGG GAGGAGCGG TCTGAGCTT TG                                292

```

(2) INFORMATION FOR SEQ ID NO:106:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 154 base pairs

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:106:

GATCGTACCC GTCCGAGTGC TCGGGCCGTT TGAGSATGGA GTGCACGTGT CTTTCGTGAT	60
GGCATACCCA GAGATGTTGG CGGCGGCGGC TGACACCCTG CAGAGCATCG GTGCTACCAC	120
TGTGGCTAGC AATGCCGCTG CGGCGGCCCC GACGACTGGG GTGGTCCCCC CCGCTGCCGA	180
TGAGGTGTCT GCGCTGACTG CGGCGCACTT CGCCGCACAT GCGGCGATGT ATCAGTCCGT	240
GAGCGCTCGG GCTGCTGCGA TTCATGACCA GTTCGTGGCC ACCCTTGCCA GCAGCGCCAG	300
CTCGTATGCG GCCACTGAAG TCGCCAATGC GCGGGCGGTC AGCTAAGCCA GGAACAGTGC	360
GCACGAGAAA GCACGAGAAA TAGGGACACG TAATGGTGA TTTGGGGGCG TTACCACCGG	420
AGATCAACTC GCGGAGGAGG TACGCGCGGC CGGTTTCGGC TTCGCTGGTG CCGCGCGCTC	480
AGATGTGGGA GAGCGTGGCG AGTGACCTGT TTTGGGCGGC GTGCGCGTTC CAGTCCGTGG	540
TCTGGGCTGT GACGCTGGGG TCGTGGATAG GTTCGTGGGC GGGTCTGATG GTGGCGCGCG	600
CCTCGCGCTA TGTGGCGTGG ATGAGCGTCA CCGCGGGGCA GCGCGAGCTG ACCGCGCGCC	660
AGCTCGGGGT TGCTCGCGCG GCTACGAGA CGCGGTATGG CCTGACGGTG CCGCGCGCGG	720
TGATCGCGCA GAACCTGCTT GAAGTGTGA TTCTGATAGC GACCAACCTC TTGGGGCAAA	780
AAACCGCGGC GATCGCGCTC AACGAGCGCG AATACGGCGA GATGTGGGCG CAAGACGCGC	840
CGCGATGTTT TCGGTACGCG CGCGCGACCG GAGCGCGAG CCGGACCTTC CTGCGCTTCG	900
AGGAGGCGCC CGACATGACC AGCGCGGCTT GCTCTCTGCA GAGCGCGCGC CGCGTGGAGC	960
AGGCTTCGCA GACCGCGCGC CGGAACCACT TATGAAGAA TGTGCGCGAG CGCGTGCAC	1020
AGCTGGCGCA GCGACCGCAG GCGACGACCG GTTCTTCGAA GCTCGGTGGC CTGTGGAAGA	1080
CGCTCTCGCG GCATCGCTCG CGCATGCGCA ACATCTCTTC GATGCGCAAG AACCATGTGT	1140
GAATGATGA TTGCGCTGCT TCGATGAGCA AGATCTCTAG TTTGATCTTC AAGCGCTTTC	1200
TTGCGCTGCT TCGCGCGCAG CGCTGCGAAA CGCTGCGGCA AAACGCGCTC CGCGTGCATG	1260
CGCTGCTGCG TACCTCGCTC GTTCTCTGCG GTCTGCGGCG TCGGCTGCGC GCGAAGCTTC	1320
GTGCGCGCGC CTGCTCTGCT TCGCTCTGCG TCGCGCGCGC TCGGCTGCGC GCGAAGCTTC	1380
GAATGATGA TCGCGCTGCT TCGCTCTGCG TCGCGCGCGC TCGGCTGCGC GCGAAGCTTC	1440

TTCCGCGGCC GGCTATGACA ACAGTCAATG TGCATGACAA GTTACAGGTA TTAGGTCCAG 1680
 GTTCAACAAG GAGACAGGCA ACATGGCCTC ACGTTTTATG ACGGATCCGC ACGCGATGCG 1740
 GGACATGGCG GGCCGTTTTG AGGTGCACGC CCAGACGGTG GAGGACGAGG CTCGCCGGAT 1800
 GTGGGCGTCC GCGCAAAACA TTTCCGGTGC GGGCTGGAGT GGCATGGCCG AGGCGACCTC 1860
 GCTAGACACC ATGGCCCAGA TGAATCAGGC GTTTCGCAAC ATCGTGAACA TGCTGCACGG 1920
 GGTGCGTGAC GGGCTGGTTC GCGACGCCAA CAACTACGAG CAGCAAGAGC AGGCCTCCCA 1980
 GCAGATCCTC AGCAGCTAAC GTCAGCCCTC GCAGCACAAT ACTTTTACAA GCGAAGGAGA 2040
 ACAGGTTTGA TGACCATCAA CTATCAATTC GGGGATGTCT ACGCTCACGG CGCCATGATC 2100
 GCGGTCAGG CCGGGTTGCT GAGGCGCGAG CATCAGGCCA TCATTGCTGA TGTGTTGACC 2160
 GCGAGTGAAT TTTGGGGCGG CGCGGTTCTG GCGGCTGCCC AGGGGTTTAT TACCCAGTTG 2220
 GGGCTTAAT TCCAGGTGAT CTACGAGCAG GCCAACGCCC ACGGGCAGAA GGTGCAGGCT 2280
 GCGGCGAACA ACATGGCGCA AACCGACAGC GCGCTCGGCT CCAGCTGGGC CTGACACCAG 2340
 GCGAAGGCCA GCGACGTGCT GTACGAGTGA AGTTCCTGCG GTGATCCTTC GGGTGGCAGT 2400
 CTAAGTGGTC AGTGCTGGG TGTGCTGGT TTGCTGCTTG GCGGTTCTT CGGTGCTGGT 2460
 CAGTGTCTCT GGGCTCGGG TGAGGACCTC GAGGCCCAGG TAGCCCGCTC CTTCGATCCA 2520
 TTCTCTCTCT TGTTCGGCA GACGCTCTC GACGAGGCG ATGATCGAGG CGCGGTGCGG 2580
 GAAGATGCCC AGGACGTGG TTCCGCTCTG TACCTCTGCG TTGAGCGCTT CCGGGGGTT 2640
 GTTGGACCAG ATTTGGCGCT AGATCTCTCT GCGGAAGGCG GTGAACGCCA GCAGGTCTGT 2700
 GCGGCGCGTG TCGAGGTGCT GGGCCACCGC GGGGAGTTTG TCGCTCAGAG CGTCGAGTAC 2760
 TCGATATAT TGGCAACAA CTGATTCTCT GTGGGCTG TGTGATCTG AGTGCAGCAG 2820
 GTTGTGCTCT GAGGCGCAG AGGTTCTCT GTGCTCTG ATGAGATTCT GTGCTATCT 2880
 GTTGTGCTCT GCGTGGCAGG GCGTGTCTCT GAGGCTCTG TGTGATCTCT GCGGAGGCG 2940
 GTTGTGCTCT TCGCTGCTGA GCGTGTCTCT GCGGAGCAG TCGGCGCGCA GCGGCTGCG 3000
 GAAGAAGCCC AGCCAGCCGG TCGTGTCTCT GCGGAGGCTG AGTGTGATCT GCGGATCT 3060

INFORMATION FOR SEQ ID NO 1:

123

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:107:

Met Val Asp Phe Gly Ala Leu Pro Pro Glu Ile Asn Ser Ala Arg Met
 1 5 10 15
 Tyr Ala Gly Pro Gly Ser Ala Ser Leu Val Ala Ala Ala Gln Met Trp
 20 25 30
 Asp Ser Val Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Gln Ser
 35 40 45
 Val Val Trp Gly Leu Thr Val Gly Ser Trp Ile Gly Ser Ser Ala Gly
 50 55 60
 Leu Met Val Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr
 65 70 75 80
 Ala Gly Gln Ala Glu Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala
 85 90 95
 Ala Tyr Glu Thr Ala Tyr Gly Leu Thr Val Pro Pro Pro Val Ile Ala
 100 105 110
 Glu Asn Arg Ala Glu Leu Met Ile Leu Ile Ala Thr Asn Leu Leu Gly
 115 120 125
 Gln Asn Thr Pro Ala Ile Ala Val Asn Glu Ala Glu Tyr Gly Glu Met
 130 135 140
 Trp Ala Gln Asp Ala Ala Ala Met Phe Gly Tyr Ala Ala Ala Thr Ala
 145 150 155 160
 Thr Ala Thr Ala Thr Leu Leu Pro Phe Glu Glu Ala Pro Glu Met Thr
 165 170 175
 Ser Ala Gly Gly Leu Leu Glu Gln Ala Ala Ala Val Glu Glu Ala Ser
 180 185 190
 Asp Thr Ala Ala Ala Asn Ala Leu Met Asn Asn Val Pro Gln Ala Leu
 195 200 205
 Gln Gln Leu Ala Val Gln Thr Gln Gly Thr Thr Pro Ser Ser Lys Leu
 210 215 220
 Gly Gly Leu Trp Lys Thr Val Ser Pro His Arg Ser Pro Ile Ser Asn
 225 230 235 240
 Met Val Ser Met Ala Asn Asn His Met Ser Met Thr Asn Ser Gln Thr
 245

124

Ala Ala Ala Gln Ala Val Gln Thr Ala Ala Gln Asn Gly Val Arg Ala
 275 280 285

Met Ser Ser Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu Gly Gly Gly
 290 295 300

Val Ala Ala Asn Leu Gly Arg Ala Ala Ser Val Gly Ser Leu Ser Val
 305 310 315 320

Pro Gln Ala Trp Ala Ala Ala Asn Gln Ala Val Thr Pro Ala Ala Arg
 325 330 335

Ala Leu Pro Leu Thr Ser Leu Thr Ser Ala Ala Glu Arg Gly Pro Gly
 340 345 350

Gln Met Leu Gly Gly Leu Pro Val Gly Gln Met Gly Ala Arg Ala Gly
 355 360 365

Gly Gly Leu Ser Gly Val Leu Arg Val Pro Pro Arg Pro Tyr Val Met
 370 375 380

Pro His Ser Pro Ala Ala Gly
 385 390

12. INFORMATION FOR SEQ ID NO:108:

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1725 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

11. SEQUENCE DESCRIPTION: SEQ ID NO:108:

TACCTCAGCA CCGGCGGTGC AGGGCTGGAG CTTGCTGCTT TTTGATCTGC GGTCAAGGTG 60

AGCTCCCTCG GCGTGTGCGC GCGCTGGATG CAGATTGAT CCGGCTCTTT AGTGCAACTA 120

ATTTCCTTGA AGTGGCTGCG AGGTATAGCA TTTCCTGAT GTTAAATGTA GCGTTCACGC 180

TTTCTGGCGT TCGATTTCGG TGGAGCACTG TTGA TTAAT TTTGGGCTGC GCGCCAGGCG 240

TTGATGAGA TCGCTTGANT AATATCAAT TTTCCTGAT TGGGCGGATG CTGGGGGTAA 300

ATGAGGAGGA GCACGCTTGT CTTTCTCTCT TGAATGAG ATGTTGGCGG CCGCGGCTGG 360

TAACTTCTT TCGCTGGGGG CAACTGTGAA GGCTAGCAAT GCGGCGCGAG CCGTGGCGAC 420

TTTCTGCTGT TCGGCGCGCG TGAATGAGA TTTCCTGAT TTTCTGCA CAGAAATCGG 480

CAATCTGAT TGAATGATG TGAATGATG

CCGGGCGGTTA CCACCGGAGA TCAACTCCGC GAGGATGTAC GCCGGCCCCG GTTCGGCCTC 720
 GCTGGTGGCC GCCCGAAGA TGTGGGACAG CGTGGCGAGT GACCTGTTTT CGGCCCGCTC 780
 GCGGTTTCAG TCGGTGGTCT GGGGTCTGAC GGTGGGGTCC TGGATAGGTT CGTCGGCGGG 840
 TCTGATGGCG GCGGCGGCCT CGCGTATGT GCGGTGGATG AGCGTCACCG CGGGGCAGGC 900
 CCAGCTGACC GCCGCCCAGG TCAGGGTTGC TCGGGCGGCC TACGAGACAG CGTATAGGCT 960
 GACGGTGCCC CCGCCGGTGA TCGCCGAGAA CCGTACCGAA CTGATGACCG TGACCGCGAC 1020
 CAACCTCTTG GGGCAAAACA CGCCGGCGAT CGAGGCCAAT CAGGCCGCAT ACAGCCAGAT 1080
 GTGGGGCCAA GACCGGAGG CGATGTATGG CTACGCCGCC ACGGCGGGCA CGGCGACCA 1140
 GCGCTTGCTG CCGTTCGAGG AGCGCCCACT GATCACCAC CCGGGCGGGC TCCTTGAGCA 1200
 GCGCGTGGCG GTGAGGAGG CCATCGACAC CGCCCGCGCG AACCAGTTGA TGAACAATGT 1260
 GCGCCAAAGG CTGCAACAGC TGGCCAGCC AGCGCAGGGC GTCGTACCTT CTTCGAAGCT 1320
 GGGTGGGCTG TGGACGGCG TCTCGCGCA TGTGTGGCG CTCAGCAACG TCAGTTCGAT 1380
 AGCCAACAAC CACATGTGCA TGATGGGCAC GGGTGTGTG ATGACCAACA CATTGCACTC 1440
 GATGTTGAAG GCGTTACCTG CCGCGCGCGC TCAGGCGCTG GAAACCGCGG CGGAAAACGG 1500
 GTGTGCGCG ATGAGCTCG TGGCAGCCA GCTGGGTTCC TCGCTGGGTT CTTCGGGTCT 1560
 GCGCGCTGG GTGGCGCCA ATTGGGTG GCGGCTGCG GTGGTTGCT TGTGGTGGC 1620
 TCGAGCTCG CCGCGCGCA AGAGCGCGT CAGGCGCGG GCGCGCGCG TCGGCTGAC 1680
 TAGGCTGACC AGCGCGCGCG AAACCGCGCG CGGACACATG GTGGC 1725

INFORMATION FOR SEQ ID NO:109

1. SEQUENCE CHARACTERISTICS
 - A. LENGTH: 194 amino acids
 - B. TYPE: amino acid
 - C. STRANDEDNESS:
 - D. TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:109

Val Val Asp Phe Gly Ala Leu Pro Pro Gly Ile Asn Ser Ala Arg Met
 1 5 10 15

126

Val Val Trp Gly Leu Thr Val Gly Ser Trp Ile Gly Ser Ser Ala Gly
 50 55 60
 Leu Met Ala Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr
 65 70 75 80
 Ala Gly Gln Ala Gln Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala
 85 90 95
 Ala Tyr Glu Thr Ala Tyr Arg Leu Thr Val Pro Pro Pro Val Ile Ala
 100 105 110
 Glu Asn Arg Thr Glu Leu Met Thr Leu Thr Ala Thr Asn Leu Leu Gly
 115 120 125
 Gln Asn Thr Pro Ala Ile Glu Ala Asn Gln Ala Ala Tyr Ser Gln Met
 130 135 140
 Trp Gly Gln Asp Ala Glu Ala Met Tyr Gly Tyr Ala Ala Thr Ala Ala
 145 150 155 160
 Thr Ala Thr Glu Ala Leu Leu Pro Phe Glu Asp Ala Pro Leu Ile Thr
 165 170 175
 Asn Pro Gly Gly Leu Leu Glu Gln Ala Val Ala Val Glu Glu Ala Ile
 180 185 190
 Asp Thr Ala Ala Ala Asn Gln Leu Met Asn Asn Val Pro Gln Ala Leu
 195 200 205
 Gln Gln Leu Ala Gln Pro Ala Gln Gly Val Val Pro Ser Ser Lys Leu
 210 215 220
 Gly Gly Leu Trp Thr Ala Val Ser Pro His Leu Ser Pro Leu Ser Asn
 225 230 235 240
 Val Ser Ser Ile Ala Asn Asn His Met Ser Met Met Gly Thr Gly Val
 245 250 255
 Ser Met Thr Asn Thr Leu His Ser Met Ser Lys Gly Leu Ala Ser Val
 260 265 270
 Ala Ala Gln Ala Val Gly Thr Ala Ala Glu Asn Gly Val Trp Ala Met
 275 280 285
 Ser Ser Leu Gly Ser Gln Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu
 290 295 300
 Leu Ala Gly Val Ala Ala Asn Leu Gly Val Ala Ala Ser Val Ile Thr
 305

Ala Ala Arg Ala Leu Pro Leu Thr Ser Leu Thr Ser Ala Ala Gln Thr
 340 345 350

Ala Pro Gly His Met Leu Gly
 355

(2) INFORMATION FOR SEQ ID NO:110:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3027 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:110:

AGTTCACTCG AGAATGATAC TGACGGGGCTG TATCCAGCAT GGCTGAGACA ACCGAACCAAT 60
 CCTCCGACCC GGGGACATCG CAAGCCGACC CGATGGGCTT GGCCGCCGAA GCCGAAGCCG 120
 CCGAAGCCGA AGCCCTGGGC GGGGGGGGGG GGGCCCGTGT CCGTGCCGCC CCGTTGAAGC 180
 GTGAGGGGCT GCGCATGGCC CGAGCCGAGG ACGAGAACST CCCCAGGAT ATGCAGACTG 240
 CGAAGACCCG GAAGACTATG ACGACTATGA CGACTATGAG GCCCGAGACC AGGAGGCCCG 300
 AGGTTGGGGA TCGTGGGAG GGGGTTGGG GGTGGGTTA CCAAGACTGT CCACGATTGC 360
 CATGGGGGGG CGAGTCTCA TCATCTGGG GTTCACGGGG CTCAGCGGAT ACATTGTGTG 420
 CGACACCAT GAGGCCACCG AACCTTACCA GCGCCCGGGG GCGTTGGGG CCGGAGCCAA 480
 CGAGGTGTT ATCAACATCA CTCCTCTGCA ATTGACCAAG GCCAAGAAG AGCTGGGGCG 540
 TGTGATCGAC AGCTGACCG CGCAATTGAG CGATGACTTG CAGCAGCGGG CAGCTGATTT 600
 CAGCAGGTT GTGGAACAGT CCAAGTGGT CAGCGAAGCG ACGGTGAACG CGACAGCCGT 660
 TCACTGATG AATGAGATT CAGCTCTGT TCTGTTGTT CGACTTCAG GGGTCAGCAA 720
 TTTTCTTGT TCAAAAGG CAAGGATG ATCTTCTT CAAGTACCG TGACTGAAGA 780
 GCGGACACAG TAAAGATTG TAAAGTTCA TTTCTTA TTGACGATGA CTTAGCTGAA 840
 TTAAGACCG AAAGCATGA CGCAATGAA TCTCTTAA TCACTCAGG CGCAGCGGAA 900
 GCGCTGATG CCGGACGTA GCGATTGAG ACGAATTG CAGCGGAATG TACCGTGCAG 960
 AAATCTTAA TCAAGTTGA TCTCTTGA ATCTTCTT TTTTAAAGT CTTCTCGGT 1020

AGGTCCGCACC TCGCCGGCGA TTCTCTGTCC TATACGACCA GTTCACGCAG CAGATCGTGG 1260
 CTCGGGGCGG CAAACAGAAG TCACTGAAAA CCACCGCCAA GGTGGTGCGC GCGGCCGTGT 1320
 CGGAGCTACA TCCGGATTCC GCCGTCGTTC TGGTTTTTGT CGACCAGAGC ACTACCAGTA 1380
 AGGACAGCCC CAATCCGTCG ATGGCGGGCA GCAGCGTGAT GGTGACCCTA GCCAAGGTCC 1440
 ACGGCAATTG GCTGATCAUC AAGTTCACCC CGGTTTAGGT TGCCGTAGGC GGTGCGCAAG 1500
 TGTGACGGGG GCGCGGGTGG CTGCTCGTGC GAGATACCGG CCGTTCTCCG GACAATCAGC 1560
 GCGGACCTC AAACAGATCT CGGCCGCTGT CTAATCGGCC GGGTTATTTA AGATTAGTTG 1620
 CCACTGTATT TACCTGATCT TCAGATTGTT CAGCTGGATT TAGCTTCCCG GCAGGGCGCG 1680
 TGGTCACTT TGCATCTGGG GTTGTGACTA CTTGAGAGAA TTGAGCTGT TCGCGACCTT 1740
 GTTGTCTGTC CATCATTTGG GCTAGTTATG GCGGAGCGGA AGGATTATCG AAGTGGTGG 1800
 CTCGGGGGCG TTACCACCGG AGATCAACTC CGCGAGGATG TACGCGGGCC CGGCTTCGGC 1860
 CTCGCTGGTG GCGCGCGCGA AGATGTGGGA CAGCGTGGCG AGTGACCTGT TTTCGGCCGC 1920
 CTCGCGCTTT CATTGGGTGG TCTGGGGTCT GAUGACGGGA TCGTGGATAG GTTCGTCCGC 1980
 TGGTCTGATG GTGGCGGCGG CCTCGCGCTA TGTGGCGTGC ATGAGCGTCA CCGCGGGGCA 2040
 TCGGAGCTC ACCGCGCGCC AGGTCCGGGT TGTGCGGCG GCGTACGAGA CGGCSTATGG 2100
 TGTGACGGTG CCGCGCGCGG TGATCGCGGA GAAGCTGCT GAAGCTGATG TTGTGATAG 2160
 TATCAACCTC TTGGGGGCAA AACGCGCGCC GATGCGGCT AACGAGCGCG AATACGGGGA 2220
 TATGTGGGCG CAAGACGCGG CCGCATCTT TGGCTAGCGG GCGACGCGCG CGACGGCGAC 2280
 TGAGCGCTTG CTCGCTTCC AGGACGCGCC ACTGATCAGG AACGCGCGCG GCGTCTTGA 2340
 TCACTGCTG GCGGTGAGG AAGCTATGGA AAGCTGCTG TCAAGCGCT TGATGAACAA 2400
 TGTGCGGAA TCGTGGAA CCACTGCGCC GATCTGTCT TCGTGGAGCA ACATCGTGT 2460
 TATCTGAA AAGCACTGT GATGAGAA CTCGCTGCT TCGATGCGCA GCACCTTGA 2520
 CTAACTTTG AAGGCTTTC CTCGCTGCT TCGTGGCGG GTGAAAGCG CCGCGGAAA 2580
 TCGCTGCA TCAATAGCT TCGTGGAA TCACTGCTG TCGTGGCTG TCGTGGCTG

GACCAGCCTG ACCAGCGCCG CCCAAACCGC CCCCAGACAC ATGCTGGGCG GGCTACCGCT 2880
 GGGGCAACTG ACCAATAGCG GCGGCGGGTT CGGCGGGGTT AGCAATGCGT TCGCGATGCC 2940
 GCGCGGGGCG TACGTAATGC CCGTGTGCC CGCGCGCGGG TAACGCCGAT CCGCACGCA 3000
 TCGGGGCCCT CTATGCGGCC AGCGATC 3027

(2) INFORMATION FOR SEQ ID NO:111:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 396 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:111:

Val Val Asp Phe Gly Ala Leu Pro Pro Glu Ile Asn Ser Ala Arg Met
 1 5 10 15
 Tyr Ala Gly Pro Gly Ser Ala Ser Leu Val Ala Ala Lys Met Trp
 20 25 30
 Asp Ser Val Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Gln Ser
 35 40 45
 Val Val Trp Gly Leu Thr Thr Gly Ser Trp Ile Gly Ser Ser Ala Gly
 50 55 60
 Leu Met Val Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr
 65 70 75 80
 Ala Gly Gln Ala Gln Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala
 85 90 95
 Ala Tyr Gln Thr Ala Tyr Gly Leu Thr Val Pro Pro Pro Val Ile Ala
 100 105 110
 His Asn Arg Ala Ile Leu Met Ile Leu Ile Ala Thr Asn Leu Leu Gln
 115 120 125
 Gln Asn Thr Pro Ala Ile Ala Val Asn Ala Ala Gln Tyr Gly Gln Met
 130 135 140
 Trp Ala Gln Asp Ala Ala Ala Met Phe Gly Tyr Ala Ala Thr Ala Ala
 145 150 155 160
 Thr Ala Thr Gln Ala Leu Leu Pro Pro Ile Asp Ala Pro Leu Ile Thr
 165 170 175 180

130

195	200	205
Gln Gln Leu Ala Gln Pro Thr Lys Ser Ile Trp Pro Phe Asp Gln Leu 210 215 220		
Ser Glu Leu Trp Lys Ala Ile Ser Pro His Leu Ser Pro Leu Ser Asn 225 230 235 240		
Ile Val Ser Met Leu Asn Asn His Val Ser Met Thr Asn Ser Gly Val 245 250 255		
Ser Met Ala Ser Thr Leu His Ser Met Leu Lys Gly Phe Ala Pro Ala 260 265 270		
Ala Ala Gln Ala Val Glu Thr Ala Ala Gln Asn Gly Val Gln Ala Met 275 280 285		
Ser Ser Leu Gly Ser Gln Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu 290 295 300		
Gly Ala Gly Val Ala Ala Asn Leu Gly Arg Ala Ala Ser Val Gly Ser 305 310 315 320		
Leu Ser Val Pro Gln Ala Trp Ala Ala Ala Asn Gln Ala Val Thr Pro 325 330 335		
Ala Ala Arg Ala Leu Pro Leu Thr Ser Leu Thr Ser Ala Ala Gln Thr 340 345 350		
Ala Pro Gly His Met Leu Gly Gly Leu Pro Leu Gly Gln Leu Thr Asn 355 360 365		
Ser Gly Gly Gly Phe Gly Gly Val Ser Asn Ala Leu Arg Met Pro Pro 370 375 380		
Arg Ala Tyr Val Met Pro Arg Val Pro Ala Ala Gly 385 390 395		

INFORMATION FOR SEC ID NO 111

AGENCY CHARACTERISTICS

LENGTH 101

B TYPE: nucleic acid

STRANDEDNESS Jangle

7. TOPOLOGY linear:

X1 SEQUENCE DESCRIPTION GNG TO NG 112

INFORMATION ATTACHED AT MEMPHIS AND MOBILE IS UNRELIABLE AT THIS TIME.

GGTACAAACC GCGTCAACAC AGGCCAAGAC CCGTGCGATG CAGGCGACGG CGCAAGCCGC 300
 GGCATACACC CAGGCCATGG CCACGACGCC GTGCGTGCCG GAGATCGCCG CCAACCACAT 360
 CACCCAGGCC GTCCTTACGG CCACCAACTT CTTGGGTATC AACACGATCC CGATCGCCTT 420
 GACCGAGATG GATTATTTC A TCGTATGTG GAACCAGGCA GCCCTGGCAA TGGAGGTCTA 480
 CCAGGCCGAG ACCGCGGTTA ACACGCTTTT CGAGAAGCTC GAGCCGATGG CGTCGATCCT 540
 TGATCCCGGC GCGAGCCAGA GCACGACGAA CCGGATCTTC GGAATGCCCT CCCCTGGCAG 600
 CTCAACACCG GTTGGCCAGT TGCCGCCGGC GGCTACCCAG ACCCTCGGCC AACTGGGTGA 660
 GATGAGCGGC CCGATGCAGC AGCTGACCCA GCGGCTGCLC CAGGTGACGT CGTTGTTTAC 720
 CCAGGTGGGC GGCACCGCCG GCGGCAACCC AGCCGACGAG GAAGCCGGCC AGATGGGCCCT 780
 GGTCCGCCAC AGTCCGCTGT CGAACCATCC GCTCGCTGGT GGATCAGGCC CCAGCGCGGG 840
 CCGCGGCCCT CTGCGCGCGG AGTCCGTACC TGGCGCAGST GGGTCGTTGA CCGGCACGCC 900
 GGTGATGTCT CAGGTGATCG AAAAGCCGGT TGCCCCCTCG GTGATGCCGG CGGCTGCTGC 960
 CCGATCGCTG GCGACCGGTG GCGCCGCTCG GGTGGGTCCG GGAGCGATGG GCGAGGCTGC 1020
 GGAATCCGCC GGTCCACCA GCGCGGCTGT GGTGCGGCCG GCACCGCTCG CCGAGGAGCG 1080
 TGAAGAAGAC GACGAGGAGC ACTGGGACGA AGAGGAGGAC TGGTGAGCTC CCGTAATGAC 1140
 AACAGACTTC CCGGCCACCT GCGCCGGAAG ACTTCCCAAC ATTTTGGCGA GGAAGGTAAA 1200
 GAGAGAAAGT AGTCCAGCAT GCGAGAGATG AAGAGCGATG CCGCTACCTT CCGGCGAGGAG 1260
 GCGATTAATT TCGAGCGGAT CTGCGCGGAC CTGAAAACCC AGATCGACCA GGTGGAGTCC 1320
 ACCGTAGGTT CATTGACGGG TCACTGGCCG GCGCGCGCGG GGACCGCCCG CCAGGCTGCG 1380
 TTGCTGAGT TCGAAGAAAT A TTAATAAG GAGAAGCAGT AACTCGACCA GATCTGAGC 1440
 AATATTTTTC AGGCTGCTCT TCAATACTTC AGGCTGAA TC AGGAGGAGCA CCAAGGCTTC 1500
 TCGTGGAAA TCGGCTTCTG AGCTGTAAT AGGAAAAGAA AGGAGCAAAA AACATGACAG 1560
 AGCAGCATG GAATTTGCGT GGTATCGAGG CCGCGGCAAG TCGAATGCAAG CGAAAT 1616

2. INFORMATION FOR SEQ ID NO. 113:

SEQUENCE CHARACTERISTICS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

CTAGTGGATG GGACCATGGC CATTCTCTGC AGTCTCACTG CTTTCTGTGT TGACATTTTG 60
 GCACGCCGGC GGAAACGAAG CACTGGGGTC GAAGAACGGC TGGGCTGCCA TATCGTCCGG 120
 AGCTTCCATA CCTTCGTGCG GCCGGAAGAG CTTGTCTAG TCGGCGGCCA TGACAACCTC 180
 TCAGAGTGCG CTCAAACGTA TAAACACGAG AAAGGGCGAG ACCGACGGAA GGTGGAACTC 240
 GCGCGATCCC GTGTTTCGCT ATTCTACGCG AACTCGGCGT TGGCCTATGC GAACATCCCA 300
 GTGACGTTGC CTTGGGTGCA AGCCATTGCC TGACCGGCTT CCGTATCGT CCGCGCCAGG 360
 TTCTGCAGCG CGTGTTCAG CTCGGTAGCC GTGGCGTCCC ATTTTGTCTG GACACCCTGG 420
 TACGCTCCG AA 480

2. INFORMATION FOR SEQ ID NO:114:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 368 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

Met Leu Trp His Ala Met Pro Pro Glu Xaa Asn Thr Ala Arg Leu Met 1 5 10 15
 Ala Gly Ala Gly Pro Ala Pro Met Leu Ala Ala Ala Ala Gly Trp Glu 20 25 30
 Thr Leu Ser Ala Ala Leu Asp Ala Gln Ala Val Glu Leu Thr Ala Arg 35 40 45
 Leu Asn Ser Leu Gly Gln Ala Trp Thr Gly Gly Gly Ser Asp Lys Ala 50 55 60
 Leu Ala Ala Ala Thr Pro Met Val Val Thr Leu Gln Thr Ala Ser Thr 65 70 75 80
 His Ala Lys Thr Arg Ala Met Gln Ala Thr Ala Gln Ala Ala Tyr 85 90 95 100
 Thr Gln Ala Met Ala Thr Thr Pro Ser Leu Pro Gln Ile Ala Ala Asn 105 110 115
 His Ile Thr Gln Ala Val Leu Thr Val Thr Asn Thr Pro 120 125 130

133

Asn Gln Ala Ala Leu Ala Met Glu Val Tyr Gln Ala Glu Thr Ala Val
 145 150 155 160
 Asn Thr Leu Phe Glu Lys Leu Glu Pro Met Ala Ser Ile Leu Asp Pro
 165 170 175
 Gly Ala Ser Gln Ser Thr Thr Asn Pro Ile Phe Gly Met Pro Ser Pro
 180 185 190
 Gly Ser Ser Thr Pro Val Gly Gln Leu Pro Pro Ala Ala Thr Gln Thr
 195 200 205
 Leu Gly Gln Leu Gly Glu Met Ser Gly Pro Met Gln Gln Leu Thr Gln
 210 215 220
 Pro Leu Gln Gln Val Thr Ser Leu Phe Ser Gln Val Gly Gly Thr Gly
 225 230 235 240
 Gly Gly Asn Pro Ala Asp Glu Gln Ala Ala Gln Met Gly Leu Leu Gly
 245 250 255
 Thr Ser Pro Leu Ser Asn His Pro Leu Ala Gly Gly Ser Gly Pro Ser
 260 265 270
 Ala Gly Ala Gly Leu Leu Arg Ala Gln Ser Leu Pro Gly Ala Gly Gly
 275 280 285
 Ser Leu Thr Arg Thr Pro Leu Met Ser Gln Leu Ile Gln Lys Pro Val
 290 295 300
 Ala Pro Ser Val Met Pro Ala Ala Ala Ala Gly Ser Ser Ala Thr Gly
 305 310 315 320
 Gly Ala Ala Pro Val Gly Ala Gly Ala Met Gly Gln Gly Ala Gln Ser
 325 330 335
 Gly Gly Ser Thr Arg Pro Gly Leu Val Ala Pro Ala Pro Leu Ala Gln
 340 345 350
 Glu Asp His Glu Asp Asp His Asp Asp Thr Asp Ile His Asp Asp Thr
 355 360 365

INFORMATION FOR SEQ ID NO: 113

1. SEQUENCE CHARACTERISTICS

- A. LENGTH: 100 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS
- D. TOPOLOGY: linear

Asn Phe Glu Arg Ile Ser Gly Asp Leu Lys Thr Gln Ile Asp Gln Val
 20 25 30
 Glu Ser Thr Ala Gly Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly
 35 40 45
 Thr Ala Ala Gln Ala Ala Val Val Arg Phe Gln Glu Ala Ala Asn Lys
 50 55 60
 Gln Lys Gln Glu Leu Asp Glu Ile Ser Thr Asn Ile Arg Gln Ala Gly
 65 70 75 80
 Val Gln Tyr Ser Arg Ala Asp Glu Glu Gln Gln Ala Leu Ser Ser
 85 90 95
 Gln Met Glu Phe
 100

(E) INFORMATION FOR SEQ ID NO:116:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 396 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:116:

GATCTCCGGG GACCTGAAAA CCCAGATCGA CCAGGTGGAG TCGACGGCAG GTTCGTTGCA 60
 GGGCCAGTGG CCGCCGCCCG CGGGGACGGG CCCCCAGGG GCGGTGGTGC CTTTCCAAGA 120
 ACCAGCCCAT AAGGAGAAAG AGGAATTGTA CGAGATCTCG ACCAATATTG CTCAGGCCCG 180
 TGTCCAATAC TCGAGGGCGG ACCAGGAGCA GCAGGAGGGG CTGTCTCTGG AAATCGGCTT 240
 CTGACCCCGT AATACGAAAA GAAACGGAGG AAAAATATGA CAGAGCAGCA GTGGAATTTC 300
 CCGGCTATCG AGGCGGAGG AAGCCCAATC AAGGAAATG TCAGCTTCAT TCATTCCGTC 360
 TTGAGGAGG CGAAGGAGTC CTGAGGAAAT TTGAGG 396

(F) INFORMATION FOR SEQ ID NO:117:

- (1) SEQUENCE CHARACTERISTICS:
- A LENGTH: 80 amino acids
 - B TYPE: amino acid
 - C STRANDEDNESS: single
 - D TOPOLOGY: linear

Gly Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln
 20 25 30
 Ala Ala Val Val Arg Phe Gln Glu Ala Ala Asn Lys Gln Lys Gln Glu
 35 40 45
 Leu Asp Glu Ile Ser Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser
 50 55 60
 Arg Ala Asp Glu Glu Gln Gln Gln Ala Leu Ser Ser Gln Met Gly Phe
 65 70 75 80

(2) INFORMATION FOR SEQ ID NO:118:

1. SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 387 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(3) SEQUENCE DESCRIPTION: SEQ ID NO:118:

GTGGATCCCG ATCCCGTGTT TCGCTATTCT ACACGAACCT GGCCTTGCCC TATGCGAACA 60
 TCCGAGTGAC GTTGCCTTCG GTCGAAGCCA TTCCCTGACC GGCCTTGCTG ATCCTCCGCG 120
 CCAGGTTCTG CAGCGGCTTG TTCAGCTCGG TAGCCGTGGC GTCCCATTTT TCCTGGACAC 180
 CCTCGTAGCG TTCCGAACCG CTACCGCCCG AGCCCGCTGC CAGCTTGCTC AGGGACTGCT 240
 TCCCTCTCTC AAGGAGGCGA TGAATGGAGG TGACATTTCC CTGGATTGCG TTTCGCCCGG 300
 CCTCGATAGC TCGCAATTC CACTGCTGCT CTCTCATGTT TTTCCTGCTT TCTTTTTCGT 360
 ATTAGCGGGT CAGAAGGCCA TTTCGCA 387

2. INFORMATION FOR SEQ ID NO:119:

1. SEQUENCE CHARACTERISTICS:
 A. LENGTH: 387 base pairs
 B. TYPE: nucleic acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

3. SEQUENCE DESCRIPTION: SEQ ID NO:119:

TTCGAGAGG ATCTGCTTTT ACCGAGGCTT TTTTGGAGT GATGATTTCT TTTTGGAGT
 TTTTGGAGT ATCTGCTTTT TTTTGGAGT TTTTGGAGT TTTTGGAGT TTTTGGAGT

GGCGGGGGTT GCGCGATTGG CATCTTGGCC CA

272

(2) INFORMATION FOR SEQ ID NO:120:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

Asp Pro Val Asp Ala Val Ile Asn Thr Thr Cys Asn Tyr Gly Gln Val
 1 5 10 15
 Val Ala Ala Leu
 20

(2) INFORMATION FOR SEQ ID NO:121:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

Ala Val Glu Ser Gly Met Leu Ala Leu Gly Thr Pro Ala Pro Ser
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:122:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 13 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

Ala Ala Met Lys Pro Ala Ile Glu Asn Gly Pro Leu Glu Ala Ala Lys
 1 5 10 15
 Glu Gly Arg

(2) INFORMATION FOR SEQ ID NO:123:

- (i) SEQUENCE CHARACTERISTICS:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

Tyr Tyr Trp Cys Pro Gly Gln Pro Phe Asp Pro Ala Trp Gly Pro
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:124:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 14 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

Asp Ile Gly Ser Glu Ser Thr Glu Asp Gln Gln Xaa Ala Val
1 5 10

(2) INFORMATION FOR SEQ ID NO:125:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 13 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

Ala Glu Glu Ser Ile Ser Thr Xaa Glu Xaa Ile Val Pro
1 5 10

(2) INFORMATION FOR SEQ ID NO:126:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 17 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

Asp Pro Glu Pro Ala Pro Pro Val Thr Thr Thr Ala Ala Ser Pro Pro
1 5 10 15

Ser

(2) INFORMATION FOR SEQ ID NO:127:

- (i) SEQUENCE CHARACTERISTICS:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

Ala Pro Lys Thr Tyr Xaa Glu Glu Leu Lys Gly Thr Asp Thr Gly
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:128:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

Asp Pro Ala Ser Ala Pro Asp Val Pro Thr Ala Ala Gln Leu Thr Ser
1 5 10 15
Leu Leu Asn Ser Leu Ala Asp Pro Asn Val Ser Phe Ala Asn
20 25 30

(2) INFORMATION FOR SEQ ID NO:129:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

Asp Pro Pro Asp Pro His Gln Xaa Asp Met Thr Lys Gly Tyr Tyr Pro
1 5 10 15
Gly Gly Arg Arg Xaa Phe
20

(2) INFORMATION FOR SEQ ID NO:130:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

Asp Asp Gly Tyr Thr Ser Gly
1 7

- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: /note= "The Second Residue Can Be Either a Pro or Thr"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

Xaa	Xaa	Gly	Phe	Thr	Gly	Pro	Gln	Phe	Tyr
1				5					10

(2) INFORMATION FOR SEQ ID NO:132:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ix) FEATURE:

(D) OTHER INFORMATION: /note= "The Third Residue Can Be Either a Gln or Leu"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

Xaa	Pro	Xaa	Val	Thr	Ala	Tyr	Ala	Gly
1			5					

(2) INFORMATION FOR SEQ ID NO:133:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

Xaa	Xaa	Gln	Gln	Leu	Pro	Phe	Leu	Arg
1								

(2) INFORMATION FOR SEQ ID NO:134:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO:135:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:135:

Ala	Gly	Asp	Thr	Xaa	Ile	Tyr	Ile	Val	Gly	Asn	Leu	Thr	Ala	Asp
1					5				10				15	

(2) INFORMATION FOR SEQ ID NO:136:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:136:

Ala	Pro	Glu	Ser	Gly	Ala	Gly	Leu	Gly	Gly	Thr	Val	Gln	Ala	Gly
1				5					10				15	

(2) INFORMATION FOR SEQ ID NO:137:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 21 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:137:

Xaa	Tyr	Ile	Ala	Tyr	Xaa	Thr	Thr	Ala	Gly	Ile	Val	Pro	Gly	Lys	Ile
1				5					10					15	

Asn	Val	His	Leu	Val
				20

(2) INFORMATION FOR SEQ ID NO:138:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 882 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

[41]

GCAACGCTGT GGTAGCTTT GGGGTGATG GTTGTGCTG GGTGGGCTG GCGTGGGCG
 60
 TCACCATCCG ACCGAGCCCG GCTTAAAAAC CGGTACAGGG ACACCAAAAC GCGCAGCCAG
 120
 GGAAGTTCAT GCGGTGTTG CCGACGCAAC AGCAGGCGGC GGTCCCGCCG CCGCGGCGCG
 180
 ATGATCCAC CGCTGGATTC CAGGGCGGCA CCATTCCGCC TGTACAGAAC GTGGTGCCGC
 240
 GCGCGGGTAC CTCACCCGGG GTGGGTGGGA CCGCGGCTTC GCTCGGCG GAAGCGCCGG
 300
 CCGTGCCCGG TGTGTGCTT GCGCGGTGC CAATCCCGGT CCGGATCATC ATTCCCGCGT
 360
 TCGCGGTTG GCAGCGTGA ATGGGACGA TCGGACCCG ACCCGCGAGC AGCGCGGTGA
 420
 CCACTCCCG CAGCAGCGG CGGACGAGT TCGGAGCAG CCGCGTGAAG AGCGCGCCAA
 480
 CCGCGCGCTG CAGCAGCGG ATGACGAGT TCGGAGCAG CCGCGTGAAG AGCGCGCCAA
 540
 CCGCGCGCTG CAGCAGCGG ATGACGAGT TCGGAGCAG CCGCGTGAAG AGCGCGCCAA
 600
 CCGCGCGCTG CAGCAGCGG ATGACGAGT TCGGAGCAG CCGCGTGAAG AGCGCGCCAA
 660
 CCGCGCGCTG CAGCAGCGG ATGACGAGT TCGGAGCAG CCGCGTGAAG AGCGCGCCAA
 720
 CCGCGCGCTG CAGCAGCGG ATGACGAGT TCGGAGCAG CCGCGTGAAG AGCGCGCCAA
 780
 CCGCGCGCTG CAGCAGCGG ATGACGAGT TCGGAGCAG CCGCGTGAAG AGCGCGCCAA
 840
 CCGCGCGCTG CAGCAGCGG ATGACGAGT TCGGAGCAG CCGCGTGAAG AGCGCGCCAA
 880

(2) INFORMATION FOR SEQ ID NO:129:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 816 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(2) MOLECULE TYPE: DNA (genomic)

(3) ORIGIN: GenBank accession number: U000000000

NOTE: SEQUENCE INFORMATION FOR SEQ ID NO:130:

NOTE: SEQUENCE INFORMATION FOR SEQ ID NO:131:

NOTE: SEQUENCE INFORMATION FOR SEQ ID NO:132:

NOTE: SEQUENCE INFORMATION FOR SEQ ID NO:133:

NOTE: SEQUENCE INFORMATION FOR SEQ ID NO:134:

142

GGAGTGGCGC GATTAGGCGA CTCACCGGCG GAAGCAGCGA AAGTACTGTC GGTACCCGAG 480
 CACTTCGAGA CGACACUACA GCACGGGGTT GTGGCGCGAC TGGGTGAAAT GGCAGCCGAT 540
 AGCGECTAGC TGTCGGCTGC GGTCAACCTC GATCATGATG TCGAGGTGAC CSTGACCGCG 600
 CCCCCGAAG GAGGCGCTGA ACTCGGCGTT GAGCGGATCG GCGATCGGTT GGGGCAGTGC 660
 CCAGGCCAAT ACGGGGATAC CGGTGTCTNA ACCCGGCGCG AGCGCAGGTT CGGTTCGCGC 720
 ACGTGTGTCG GGTGCGCTTC TTACGCGCTT GTGTCGAAAC ACGAGTAGCA GGTCTGCTCC 780
 GCGGAGGCGA TCGAGGAGG GTTCTGTCAG CTCCT 815

(12) INFORMATION FOR SEQ ID NO:140:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1150 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(13) SEQUENCE DESCRIPTION: SEQ ID NO:140:

ATTAGGCGCG GGTGAGGTC TCAGATCAGA GATCTGCG AGTACCTGCG GCGCTTCAGC 50
 CTCTGCGCGA AACAGCTGCT GAAGATGCTC GCGCGCGAAA GAGGTGCGA TTTCAGGCTC 100
 TATGAGCGGT TGAACGACGA GATGATGCGG GAGATGATA TCGAAGCTT GCGGTAAACAG 150
 GTGCGGAAAG TGTGCGGCTT TTGCTGCTCG GAGTACGCTT AGGATGAG GTTTCGCGAC 200
 GCGCGGCTGA TGTGCTGCA GAGGAGCTT TCGAAGCGC AGTATGCTT GAGAGCGCTG 250
 GACGCGATCA GGTCTGCTT GCGCGCGCGA ATGAGGAGG GGTATGCTT TCGAGCTGCT 300
 GAGGTGCGA GAGGATGCA GTGCTGCGA GAGGAGGAG GAGGAGGAG GAGGAGGAG 350
 GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG 400
 GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG 450
 GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG 500
 GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG 550
 GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG 600
 GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG 650
 GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG 700
 GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG 750
 GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG 800
 GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG 850
 GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG 900
 GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG 950
 GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG 1000
 GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG 1050
 GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG 1100
 GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG GAGGAGGAG 1150

```

TTCCTACAGC AGGATTCAGC AGGCGACGCA AGCTCTGGT; CCGAATTTC; CAGGAGCTCC      900
CTTGAAACCG CTGGCCCGCT GCTCACTGCC CTACGTAAT CCGCTGCCCT CAGGCCGACC      960
CGCGGGCCGA ATACCAGCAG ATCGGACAGC GAATTGCCGC CCAGCCGGTC GGAGCCGTCG      1020
ATACGCCCGG CACACTCACC GGCAGCGAAC AGGCTGGCA CCGTGGCGG GCGGTGTCC      1080
CGGTCTACTT CGACACCGCC CATCACGTAG TGACACGTG CCCCAGCTTC CATTGCGTCC      1140
GTTGGGCACG AG                                         1152

```

(12) INFORMATION FOR SEQ ID NO:141:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 656 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(13) SEQUENCE DESCRIPTION: SEQ ID NO:141:

```

CTCTGCCGA TTGGGAGSS TGTACTTCCG GGTCTGTAT GCGGCATGAG TCCGACGAC      60
CAGGATGCC CGACAGGAC GGTGCGCT CAGGATTC ATTGGTCA CTGCGCAT      120
CGGTGAGT CCGCCCTGCG GGTCTTTT TTGGGCGAG GTGATCGAT GTGTTCGG      180
CTGTTTTC CCGATTTC GCGGCGCGC GGTGCGGCG GGTCTTTC TCGAAGTCC      240
ATGCTGACG CAGGATTC GGTCTTTTCA AGCTTTTGA GGTCTGCG CAGGATTC      300
CGGTGATTC CTACTAG A GATTTTCG CAGTACG A TCGAAGCA TCGGTTCG      360
TTCATCGAG GTTCATTCG GATGCGGCG AGGATGCG CTGACGAT GTGTTCGG      420
TTTGGAGG TACGATCA GTTTTTCG TGTCTTTC GTTCTTTC TTTGAGG      480
TTTCTTTC TTTTTC TTTTTCG A TTTTTC TTTTTC TTTTTC      540
TTTCTTTC TTTTTC TTTTTC TTTTTC TTTTTC TTTTTC      600
TTTCTTTC TTTTTC TTTTTC TTTTTC TTTTTC TTTTTC      660

```

(12) INFORMATION FOR SEQ ID NO:142:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 656 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:142

```

Asn Ala Val Val Ala Phe Ala Val Ile Gly Phe Ala Ser Leu Ala Val
1           5           10           15
Ala Val Ala Val Thr Ile Arg Pro Thr Ala Ala Ser Lys Pro Val Glu
20           25           30
Gly His Gln Asn Ala Gln Pro Gly Lys Phe Met Pro Leu Leu Pro Thr
35           40           45
Gln Gln Gln Ala Pro Val Pro Pro Pro Pro Asp Asp Pro Thr Ala
50           55           60
Gly Phe Gln Gly Gly Thr Ile Pro Ala Val Gln Asp Val Val Pro Arg
65           70           75           80
Pro Gly Thr Ser Pro Gly Val Gly Gly Thr Pro Ala Ser Pro Ala Pro
85           90           95
Gln Ala Pro Ala Val Pro Gly Val Val Pro Ala Pro Val Pro Ile Pro
100          105          110
Val Pro Ile Ile Ile Pro His Phe Pro Gly Trp Gln Pro Gly Met Pro
115          120          125
Thr Ile Pro Thr Ala Pro Pro Thr Thr Pro Val Thr Thr Ser Ala Thr
130          135          140
Thr Pro Pro Thr Thr Pro Pro Thr Thr Pro Val Thr Thr Pro Pro Thr
145          150          155          160
Thr Pro Pro Thr Thr Pro Val Thr Thr Thr Thr Thr Thr Pro Pro Thr
165          170          175
Thr Pro Val Thr Thr Pro Pro Thr Thr Val Ala Pro Thr Thr Val Ala
180          185          190
Pro Thr Thr Val Ala Pro Thr Thr Val Ala Pro Thr Thr Val Ala Pro
195          200          205          210
Ala Pro Ala Pro Thr Thr Thr Val Ala Pro Thr Thr Thr Thr Thr Thr
215          220          225          230
Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
235          240          245          250

```

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 174 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

```

Ile Asn Gln Pro Leu Ala Pro Pro Ala Pro Pro Asp Pro Pro Ser Pro
1           5           10           15
Pro Arg Pro Pro Val Pro Pro Val Pro Pro Leu Pro Pro Ser Pro Pro
20           25           30
Ser Pro Pro Thr Gly Trp Val Pro Arg Ala Leu Leu Pro Pro Trp Leu
35           40           45
Ala Gly Thr Pro Pro Ala Pro Pro Val Pro Pro Met Ala Pro Leu Pro
50           55           60
Pro Ala Ala Pro Leu Pro Pro Leu Pro Pro Leu Pro Pro Leu Pro TH.
65           70           75           80
Ser His Pro Pro Arg Pro Pro Ala Pro Pro Ala Pro Pro Ala Pro Pro
85           90           95
Ala Tyr Pro Phe Val Pro Val Pro Pro Ala Pro Pro Leu Pro Pro Ser
100          105          110
Pro Pro Thr Thr Leu Pro Ala Asp Ala Ala Tyr Pro Pro Ala Pro Pro
115          120          125
Ala Pro Pro Leu Ala Pro Pro Ser Pro Pro Ala Gly Ser Ala Ala Thr
130          135          140
Asp Ala Leu Thr Gly Ala Thr Ser Val Thr Thr Thr Thr Thr Thr Thr
145          150          155          160          165          170
Leu Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
175          180          185          190          195          200

```

2. INFORMATION FOR SEQ ID NO: 144:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 174 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

146

1 5 10 15
 20 25 30
 35
 40

(2) INFORMATION FOR SEQ ID NO:145:

1. SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 104 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

2. MOLECULE TYPE: peptide

(X) SEQUENCE DESCRIPTION: SEQ ID NO:145:

1 5 10 15
 20 25 30
 35 40 45
 50 55 60
 65 70 75 80
 85 90 95

(2) INFORMATION FOR SEQ ID NO:146:

1. SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 81 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

147

(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

GGATCCATAT GGGCCATAT CATCATCATC ACGTGATCGA CATCATCGGG ACC

53

(2) INFORMATION FOR SEQ ID NO:147:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: (desc) - "PCR Primer"

(iii) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

CCGCAATTCA GGGCTCGCTT GCGCCGGCCT CATCTGAAC GA

42

(2) INFORMATION FOR SEQ ID NO:148:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: (desc) - "PCR Primer"

(iii) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:148:

GATGATTA GGTGAAG CAGGATGTT

(2) INFORMATION FOR SEQ ID NO:149:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:149:

CTCTGAATTC AGCGGTGGAA ATCGTCGCGA T

31

(2) INFORMATION FOR SEQ ID NO:150:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: (desc) "PCR primer"

(iii) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:150:

GGATCCAGCG CTGAGATGAA GAGGATGCG GT

33

(2) INFORMATION FOR SEQ ID NO:151:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: (desc) "PCR primer"

(iii) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:151:

GAAGCAATT TTA GAGGATGTTTCTGAGG AAT

(2) INFORMATION FOR SEQ ID NO:152:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE:

149

(A) NAME/KEY: CD1

(B) LOCATION: 152..1523

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:152.

TGTTCCTTCGA CGGCAGGCTG GTGSAGSAAG GCGCCACCGA ACAGCTGTTC TCCTCGCCGA 60
 AGCATGCGGA AACCGGCCGA TACGTGCGCG GACTGTGCGG GGACGTCAAG GACGCCAAGC 120
 GCGGAAATTG AAGAGACAG AAAGSTATGS C GTG AAA ATT CGT TTG CAT ACG 172
 Val Lys Ile Arg Leu His Thr
 1 5
 CTG TTT GCG GTG TTG ACC GGT GAT TTT CTG CTG CTA GCA GCG GCG GCG 220
 Leu Leu Ala Val Leu Thr Ala Ala Pro Leu Leu Leu Ala Ala Ala Gly
 10 15 20
 TGT GCG TCG AAA GCA CTG AGG GGT TTT GGT GAA ACG GCG GCG GCG GCG 262
 Cys Gly Ser Lys Pro Pro Ser Gly Ser Pro Glu Thr Gly Ala Gly Ala
 25 30 35
 GGT ACT CTC GCG ACT ACC GGT GAT TCG TCG CCG GTS ACG TTG GCG GAG 318
 Gly Thr Val Ala Thr Thr Pro Ala Ser Ser Pro Val Thr Leu Ala Glu
 40 45 50 55
 ACC GGT AGG ACG CTC CTC TAT CCA CCA TTT TTT AAT CTC TCG GGT CCG GCG 362
 Thr Gly Ser Thr Leu Leu Tyr Phe Leu Phe Asn Leu Trp Gly Pro Ala
 60 65 70
 TTT GAC GAG AGG TAT CCA AAT GTT ACG ATC ACG GGT GAG GCG AGG GGT 418
 Phe His Glu Arg Tyr Pro Asn Val Thr Ile Thr Ala Gln Gly Thr Gly
 75 80 85
 TTT GGT GCG GGT ATT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT 462
 Ser Gly Ala Gly Ile Ala Gln Ala Ala Ala Gly Thr Val Asn Ile Gly
 90 95 100
 TTT TTT GAT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT 508
 Ala Ser Asp Ala Tyr Leu Ser Leu Ile Asn Met Ala Ala His Lys Gly
 105 110 115
 TTT TTT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT 554
 Leu Met Asp Ile Ala Val Ala Val Gln Ile Thr Val Val Val Val Val
 120 125 130
 TTT TTT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT 600
 Leu Met Asp Ile Ala Val Ala Val Gln Ile Thr Val Val Val Val Val
 135 140 145
 TTT TTT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT 646
 Leu Met Asp Ile Ala Val Ala Val Gln Ile Thr Val Val Val Val Val
 150 155 160

150

170	175	180	
CAC CCG TCC GAC GGG TCC GGT GAC ACC TTC TTG TTC ACC CAG TAC CTG			748
His Arg Ser Asp Gly Ser Gly Asp Thr Phe Leu Phe Thr Gln Tyr Leu			
185	190	195	
TCC AAG CAA GAT CCC GAG GGC TGG GGC AAG TCC CCC GGC TTC GGC ACC			796
Ser Lys Gln Asp Pro Glu Gly Trp Gly Lys Ser Pro Gly Phe Gly Thr			
200	205	210	215
ACC CTC GAC TTC CCG GCG GTG CCG GGT GCG CTG GGT GAG AAC GCG AAC			844
Thr Val Asp Phe Pro Ala Val Pro Gly Ala Leu Gly Glu Asn Gly Asn			
220	225	230	
GGT GGC ATG GTG ACC GGT TGT GTC GAG AAA TCG GCG TCC GTG GCG TAT			892
Gly Gly Met Val Thr Gly Cys Ala Glu Thr Pro Gly Lys Val Ala Tyr			
235	240	245	
ATC GGC ATC ACC TTC CTC GAC CAG GCG ATT CAA CCG GGA CTC GCG GAG			940
Ile Gly Ile Ser Phe Leu Asp Gln Ala Ser Gln Arg Gly Leu Gly Glu			
250	255	260	
GCG CAA CTA GGC AAT AGC TCT GCG AAT TTC TTG TCG CCG GAC GCG CAA			988
Ala Gln Leu Gly Asn Ser Ser Gly Asn Phe Leu Leu Pro Asp Ala Gln			
265	270	275	
AGC ATT CAG GCG GCG GCG GGT GCG TTC GCA TCG AAA ACC GCG GCG AAT			1036
Ser Ile Gln Ala Ala Ala Ala Gly Phe Ala Ser Lys Thr Pro Ala Asn			
280	285	290	295
TAC GCG ATT TCG ATG ATT TAC GCG GCG GCG GCG GCG GCG TAC CCG ATC			1084
His Ala Ile Ser Met Ile Asp Ile Pro Ala Pro Asn Gly Tyr Pro Ile			
300	305	310	
ATT AAC TAC GAG TAC GCG ATT GTC GCG AAC GCG CAA AAG GAC GCG GCG			1132
Ile Asn Tyr Gln Tyr Ala Ile Val Asn Asn Arg Gln Lys Asp Ala Ala			
315	320	325	
ATC GCG TAC GAG TAC GCG ATT GTC GCG AAC GCG CAA AAG GAC GCG GCG			1180
Ile Asn Tyr Gln Tyr Ala Ile Val Asn Asn Arg Gln Lys Asp Ala Ala			
330	335	340	
ATC GCG TAC GAG TAC GCG ATT GTC GCG AAC GCG CAA AAG GAC GCG GCG			1228
Ile Asn Tyr Gln Tyr Ala Ile Val Asn Asn Arg Gln Lys Asp Ala Ala			
345	350	355	
ATC GCG TAC GAG TAC GCG ATT GTC GCG AAC GCG CAA AAG GAC GCG GCG			1276
Ile Asn Tyr Gln Tyr Ala Ile Val Asn Asn Arg Gln Lys Asp Ala Ala			
360	365	370	
ATC GCG TAC GAG TAC GCG ATT GTC GCG AAC GCG CAA AAG GAC GCG GCG			1324
Ile Asn Tyr Gln Tyr Ala Ile Val Asn Asn Arg Gln Lys Asp Ala Ala			
375	380	385	
ATC GCG TAC GAG TAC GCG ATT GTC GCG AAC GCG CAA AAG GAC GCG GCG			1372
Ile Asn Tyr Gln Tyr Ala Ile Val Asn Asn Arg Gln Lys Asp Ala Ala			
390	395	400	

(2) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 374 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(00) SEQUENCE DESCRIPTION: SEQ ID NO:153

[illegible]

152

130 135 140
 Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile Lys Thr
 145 150 155 160
 Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn Leu Pro
 165 170 175
 Gly Thr Ala Val Val Pro Leu His Arg Ser Asp Gly Ser Gly Asp Thr
 180 185 190
 Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Glu Gly Trp Gly
 195 200 205
 Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val Pro Gly
 210 215 220
 Ala Leu Gly His Asn Gly Asn Gly Gly Met Val Thr Gly Cys Ala Glu
 225 230 235 240
 Thr Pro Gly Cys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp Gln Ala
 245 250 255
 Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser Gly Asn
 260 265 270
 Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala Gly Phe
 275 280 285
 Ala Ser Lys Thr Pro Ala Asn His Ala Ile Ser Met Ile Asp Gly Pro
 290 295 300
 Ala Phe Asn Gly Tyr Pro Ile Ile Asn Tyr His Tyr Ala Ile Val Asn
 305 310 315 320
 Asn Arg His Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala Phe Leu
 325 330 335
 His Thr Ala Ile Thr Asn Gly Asn Ala Ala Asn Ile Thr Val Gln Val
 340 345 350
 His Leu Ala Ile Asn Ser Ile Ala Thr Val Ala Asn Ser Asn Ala Leu
 355 360 365
 Leu Ala His Ile Ser Ser
 370

INFORMATION FOR SEQ. NO. 152

The sequence of the amino acid sequence of the protein is as follows:
 1-370

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:154:

TGTTCCTCGA CGGCAGGCTG GTGGAGGAAG GGCCCAACCGA ACAGCTGTTT TCCTCGCGGA 60
 AGCATGCGGA AACCGCCCGA TACCTCSCCG GACTGTGCGG GGACGTCAAG GACGCCAAGC 120
 GCGGAAATTG AAGAGCACAG AAAGGTATCG CGTGAAATT CGTTTGCATA CGCTGTTGCG 180
 CGTGTGAGCC GTTGGCCCGC TCCTGCTAGC AGCGGCGGGC TGTGGGTGGA AACCACCGAG 240
 AGTTTGGCTT GAAACGCGCG CGCGCGCGCG TACTGTGCGG ACTACCCCG CGTCGTGCGC 300
 GGTGACCTTG GCGGAGACCG GTAGCAGCTT GCTGTACCGG GTGTTCAAGT TGTGGGCTCG 360
 GGGTTTTCAG GAGAGGTATG TAAAGTCAAG GATCACCGCT CAGGCGACCG GTTGTGCTCG 420
 CGGATGCGCG CAGCTGCTCG TCGGACGCTT CAGCTTGGCG GCGTGGAGCG GGTATGCTCG 480
 GGAAGGTGAT ATGGCGCGCG ACAGCGCGCT GATGAACATC GCGGTAGCGA TGTGCGCTGA 540
 GTAGGTCAAG TACAGCTGCG CGGATGAG CGAGCACCTC AAGCTGACCG GAAAGCTGCT 600
 GCGCGCGCAT TACAGCGCGA GCATGAAAGC GTGGGACGAG CGCGAGATCG CTCGCGCTGA 660
 GCGCGCGCTG AAGCTGCGCG GCAGCGCGCT AGTTGCGCTG TACGCTGCGG ACGGCTGCGG 720
 TGACACCTTG TTGTTGACCG AGTACCTGTC CAAGCAAGAT GTGAGGCTT GCGGAACTT 780
 GCGCGCGCTG GCGACGACCG TCGACTTGGC GCGCGTGGCG GTGCGGCTTG GTGAGAAGC 840
 CAACGCGCGG ATCGTACGCG GTTGGCGCGA GACACGCGCG TCGCTGCGCT ATATGCTGAT 900
 CAGCTTCTCT TACGAGCGCG GTGAGCGCG AGTGGCGCGA GCGGAGTAC GAAATAGCT 960
 TCGGAACTTG TTGTTGCGCG AGTGGCGCG GATCGAGCTT GCGGCGCTG GCGGAGTAC 1020
 GAAAGCGCTT GCGGAGCGCG GCGTTGCTAT GATCGAGCTT GCGGCGCTG GCGGAGTAC 1080
 TATGATGAG TAAGGAGTAT GCGTCTGTA TAAAGGAGG AAGGAGTAT TAAAGGAGT 1140
 TAAGGAGTAT TAAAGGAGT TAAAGGAGT TAAAGGAGT TAAAGGAGT TAAAGGAGT 1200
 TAAAGGAGT TAAAGGAGT TAAAGGAGT TAAAGGAGT TAAAGGAGT TAAAGGAGT 1260
 TAAAGGAGT TAAAGGAGT TAAAGGAGT TAAAGGAGT TAAAGGAGT TAAAGGAGT 1320
 TAAAGGAGT TAAAGGAGT TAAAGGAGT TAAAGGAGT TAAAGGAGT TAAAGGAGT 1380
 TAAAGGAGT TAAAGGAGT TAAAGGAGT TAAAGGAGT TAAAGGAGT TAAAGGAGT 1440
 TAAAGGAGT TAAAGGAGT TAAAGGAGT TAAAGGAGT TAAAGGAGT TAAAGGAGT 1500

TCGCCCTGAT CATCCGCGTG CCGGTGTCTG TAGGAGCGGG GCTGTGTGATG GTGGAACGGC 1680
 TGGCGAAACG GTTGGCCGAG GCTGTGCGAA TACTCCTGGA ATTGCTCGGC GGAATCCCCA 1740
 GCGTGGTGGT CCGTTTGTGG GGGGCAATGA CGTTCGGGCG GTTCATCGCT CATCACATCG 1800
 CTCGGGTGAT CGCTCACAAC GGTCCCGATG TCGCGTGCTT GAACTACTTG CGCGGCCACC 1860
 CGGGCAACGG GGAGGGCATG TTGGTATCCG GTCTGTGTTT GCGCGTGATG GTCGTTCCCA 1920
 TTATCGCCAC CACCAATCAT GACCTTTTTC GCGAGTGTGG GGTGTTCGCG CGGGAGGGCG 1980
 CGATCGGGAA TTC 1993

2. INFORMATION FOR SEQ ID NO:155:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 374 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:155:

Met Lys Ile Arg Leu His Thr Leu Leu Ala Val Leu Thr Ala Ala Pro
 1 5 10 15
 Leu Leu Leu Ala Ala Ala Gly Tyr Gly Ser Lys Pro Pro Ser Gly Ser
 20 25 30
 Pro Glu Thr Gly Ala Val Ala Gly Thr Val Ala Thr Thr Pro Ala Ser
 35 40 45
 Ser Pro Val Thr Leu Ala Ile Thr His Leu Thr Leu Leu Tyr Pro Leu
 50 55 60
 Phe Ala Leu Thr Val Leu Ala Leu Met Val Ala Thr Leu Asn Val Thr
 65 70 75
 Thr Val Leu Val Leu Val Leu Val Leu Val Leu Val Leu Val Leu Val
 80 85 90 95 100 105 110 115
 Ala Thr Thr Val Leu Leu Val Ala Thr Leu Val Leu Leu Leu Leu Gly
 120 125 130 135 140 145 150
 Arg Thr Ala Ala His Gly Ile Ser Met Arg Thr Thr Leu Ala Leu Leu

155

Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn Leu Pro
165 170 175

Gly Thr Ala Val Val Pro Leu His Arg Ser Asp Gly Ser Gly Asp Thr
180 185 190

Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Glu Gly Trp Gly
195 200 205

Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val Pro Gly
210 215 220

Ala Leu Gly Glu Asn Gly Asn Gly Gly Met Val Thr Gly Cys Ala Glu
225 230 235 240

Thr Pro Gly Cys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp Gln Ala
245 250 255

Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser Gly Asn
260 265 270

Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala Gly Phe
275 280 285

Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp Gly Pro
290 295 300

Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile Val Asn
305 310 315 320

Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala Phe Leu
325 330 335

His Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp Gln Val
340 345 350

His Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp Ala Leu
355 360 365

1. Ala Thr Ile Ser Ser
370

SEQUENCE INFORMATION

SEQUENCE CHARACTERISTICS
 A: LENGTH 370 amino acids
 B: TYPE single chain
 C: TRANSDUCED NO. 1
 D: NO. OF CHAINS 1

GTAGCGAAAC GSCAACGGGG CCGCGTGGG CAGCTTCTTC AGCGTGATT TGCAGAACAC 180
 CTCGTGGAAG GTGATGCCGT CGAATTGTGG CCGCGAACG CTGGGGACCA GGCGGATCGG 240
 CTGCAACCCG GCAGCGCCCG TCGTCAACGG GCATCCCGTT CACCGCGACG GCTTGCCGGG 300
 CCCAACGGAT ACCATTATTC GAACAACGGT TGTATCTTT GTCAACGCTG GCCGTACCG 360
 AGCGCGCAC AGGATGTGAT ATGCCATCTG TCGCCGACA GACAGGAGCG AGGCCTTATG 420
 ACAGGATTCG GCGTCGAGCC CTACGGGAG CCGAAGTAG TAGAATGCG CCGGAAGCGG 480
 ATGGCGTATA TCGACGAAG CAAGCGTAG GATCTCTCT TCGAGGAGG CAACCGCACG 540
 TGTCTTATCT TGTGCGGAA GATCATCTCG CATTGAGAG GCGTGGGCG GCTGCTGGCG 600
 TCGCATCTGA TCGGATGGG CCGCTCGAG AAGTCAAGG CATCGGAGC CGACCGCTAT 660
 AGCTATGGG AGCAACGAG CTTTTCTTC GAGCTTGGG ATGCGCTCGA CCGCGCGAG 720
 CAGCTGCTAC TGGTGTGCA CGACTGGGG TCGCGCTCG GCTTCGACTG GGCTAACCAG 780
 CATCGGAGG GAGTGGAGG GATCGCTTC ATGAGAGCA TCGTCAGCG GATGACGTGG 840
 CGGACGCGG CCGCGCGCT TCGGCTCTT TCGAGCTT TCGGATCGG TCAAGCGGAG 900
 CGAATGGCT TGGAGCAAA CATCTCTCT GATCTCTT TCTTCTCTT GATCTCTCT 960
 CAGCTCAGG ACGAGGAAT GAGCAATAT CCGCGCGAT TCGTGAAGG CCGCGAGGAG 1020
 CCGCGCGCA CTTCTCTCT GAGAGGAAG CTTCAATAG AGCTTAGCG CTCTTAGCT 1080
 TCGCTCTCT TCAAGGAGT CCGGATCTT TCGAGGAGT TCGAGGAGT TCGAGGAGT 1140
 ATGAGGAGG AGCGCGGCG CATCTCTCT TCGGATCTT TCGGATCTT TCGGATCTT 1200
 TCGAGGAGT CCGGATCTT TCGGATCTT TCGGATCTT TCGGATCTT TCGGATCTT 1260
 TCGGATCTT TCGGATCTT TCGGATCTT TCGGATCTT TCGGATCTT TCGGATCTT 1320
 TCGGATCTT TCGGATCTT TCGGATCTT TCGGATCTT TCGGATCTT TCGGATCTT 1380
 TCGGATCTT TCGGATCTT TCGGATCTT TCGGATCTT TCGGATCTT TCGGATCTT 1440
 TCGGATCTT TCGGATCTT TCGGATCTT TCGGATCTT TCGGATCTT TCGGATCTT 1500
 TCGGATCTT TCGGATCTT TCGGATCTT TCGGATCTT TCGGATCTT TCGGATCTT 1560

TACGTGCGCG TTCACGATTC GGGGGGGGGG TATGCTT

1177

(2) INFORMATION FOR SEQ ID NO:157:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 324 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:157:

GAGATTGAAT GGTACCGGTC TCGTTAGCGG CTCGGTCCCG TGAATGAGCA TATCAGCGAC	60
GGCATGTTT TCGTTGTGCA GGTTCGGTCC ATGCTCGGAC GTTGGTAAAT TCAGGGTTTG	120
ATCAGTAATT GCGGGGACG GTTCGGGAA GCGGGGAGG ATGTGCTGA GCGGGGCGG	180
AGGTTGCGG GAGGCGAGG GTGATGCTC AGTGGCGTC GGTGAGTTA GCGAGCGTTT	240
GGGCTGCTC GTGACAGTG GTACTCCGGT GAGGAGCGG GCGGCTCGT GGTGAGAGC	300
GCTGACCGAC GCGGCGGATT CAGA	324

(2) INFORMATION FOR SEQ ID NO:158:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1338 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(2) SEQUENCE DESCRIPTION: SEQ ID NO:158:

ACCTAAGG GCGTTTGGT TGAAGGCA GGTGTAAT GTGACGAGT TCGCTCGCG	60
AAGGATTGAG GAGCGGCTG TCGGCGCT GTGCTGCTC AAGGTATGG GTGATGTCT	120
TAATATTA GCGGAGT GGTGTAAT GTGTAAT GTGCTGCTC GTGCTGCTC	180
TAATATTA GCGGAGT GGTGTAAT GTGTAAT GTGCTGCTC GTGCTGCTC	240
TAATATTA GCGGAGT GGTGTAAT GTGTAAT GTGCTGCTC GTGCTGCTC	300
TAATATTA GCGGAGT GGTGTAAT GTGTAAT GTGCTGCTC GTGCTGCTC	360
TAATATTA GCGGAGT GGTGTAAT GTGTAAT GTGCTGCTC GTGCTGCTC	420
TAATATTA GCGGAGT GGTGTAAT GTGTAAT GTGCTGCTC GTGCTGCTC	480

GGAGCCGCGG GGTGACGGG CTGCTGGAGG GCGGCGAGAG GGTGCTGCTG CGTAAGGGCG	720
GGATCGGGGA GAAGCGCTTC GAGGTGGCGG CCCACGAGTT CTTGTTGTTT CCGACGGTCG	780
CGCACAGCCA CCGCGAGCGG GTTCGCCCCG AGCACCAGCA CCTGCTGGGC CCGGCGGCCG	840
CCGACAGCAC CGACGAGTGT GTGCTACTGC GGGCCGCGAG GAAAGTTGTT GCCGCACTGC	900
CGGTTAACCG GCGGAGGGT CTGGACGCCA TCGAGGATCT GCACATCTGG ACCGCGGAGT	960
CGGTGCGGCG CGACGAGTTC GACTTTGGG CCAAGCACA ACTGGCCGTC TTGTTGGTCT	1020
CGGCGATGCG GGTGCTGAG CCGCTGCGT TGGCGCTAG GCGCGAGTAC GCGGTTGCA	1080
CGAGCTGGGT CGAGTGGCG GTGACGCGA CGTTGCGCG CCGCTGCGA GAGGAGCGCG	1140
CGCTGCGCGA GGTGCTGAG CCGCTGCGT AGGCTGCGT TCGCTGCGT GCGCTGCGT	1200
CGGTGCGAGT TGTACGCGA GTGCGGCTG CGAGTATCT GGTGCGCTT CCGCTGCGT	1260
TGCGTCAAT TGACGCGCG GCGACAGCA GATGCGCG CCGCTGCGT CCGCGCGCG	1320
GCGCGAGCG GTACAGCG	1338

(2) INFORMATION FOR SEQ ID NO:159:

(A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 321 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(B) ORIGIN: DESCRIPTION, SEQ ID NO:159:

CGGTGCGAGT GGTGCTGAG GGTGCTGAG GGTGCTGAG GGTGCTGAG GGTGCTGAG	1
TGGCTGCGT GGTGCTGAG GGTGCTGAG GGTGCTGAG GGTGCTGAG GGTGCTGAG	11
TGCTGCGT GGTGCTGAG GGTGCTGAG GGTGCTGAG GGTGCTGAG GGTGCTGAG	21
TGCTGCGT GGTGCTGAG GGTGCTGAG GGTGCTGAG GGTGCTGAG GGTGCTGAG	31
TGCTGCGT GGTGCTGAG GGTGCTGAG GGTGCTGAG GGTGCTGAG GGTGCTGAG	41
TGCTGCGT GGTGCTGAG GGTGCTGAG GGTGCTGAG GGTGCTGAG GGTGCTGAG	51
TGCTGCGT GGTGCTGAG GGTGCTGAG GGTGCTGAG GGTGCTGAG GGTGCTGAG	61

(C) INFORMATION FOR SEQ ID NO:160:

(A) SEQUENCE CHARACTERISTICS:

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:160:

GAAGACCCCGG CCCCCCATA TGGATCGGTT CGCGGACTAC TTTCGGCGAA CGTGCACGGG 60
 GCGGCGTCCG GCTGATCATC ACGGGTGGCT ACGCGCCCAA CGGCACCGGA TGGCTGCTGC 120
 CSTTCGCCTC CGAACTCGTC ACTTCGGGGC AAGCCCCGACG GCACCGCCGA ATCACCAGGG 180
 CGGTCCACGA TTCGGGTGCA AAGATCCTGC TGCAATCCT GCACGCGGA CGGTACCCCT 240
 ACCACCCACT TGCGTCAGC GCCTCGCGGA TCAAGGGGCG GATCACCCTT TTTCGTCCCG 300
 GAGCACTATC GCTCGCGGG GTGAGCGGA CCATCGCGGA TTTCGGCGG TGGGGGCACT 360
 TCGCGCGCGA TCGCGGTAC GAGGGGTG AAATCATGGG CAGCGAAGG TATCTGCTCA 420
 ATCACTTCCT GCGCGCGCGG ACAAACAGG GCACCGACTC TCGGGTGGG AAGCGCGCA 480
 ACCGTGCGCG GT 492

(X2) INFORMATION FOR SEQ ID NO:161:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 536 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(X3) SEQUENCE DESCRIPTION: SEQ ID NO:161:

Phe Ala Gln His Leu Val Ala Gly Asp Ala Val Gln Leu Trp Arg Ala
 1 5 10 15
 Asn Ala Ala Asp Gln Ala Asp Arg Leu Gln Pro Gly Ser Ala Arg Arg
 20 25 30
 Gln Arg Ala Ser Ala Ser Pro Arg Arg Leu Ala Gly His Asn Ala Tyr
 35 40 45
 His Thr Ser Asn Asn Arg Arg Leu Leu Val His Ala Thr Trp Leu Trp
 50 55
 Ser Ala Ala Leu Arg Val Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg
 60 65 70 75 80 85 90 95
 Phe Gly Leu Val Thr Ala Thr Gly Leu Thr Thr Thr Thr Thr Thr Thr Thr
 100 105 110 115 120 125 130 135 140 145
 Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
 150 155 160 165 170 175 180 185 190 195

160

130	135	140
Cys Asp Leu Ile Gly Met Gly Ala Ser Asp Lys Leu Ser Pro Ser Gly		
145	150	155 160
Pro Asp Arg Tyr Ser Tyr Gly Glu Gln Arg Asp Phe Leu Phe Ala Leu		
165	170	175
Trp Asp Ala Leu Asp Leu Gly Asp His Val Val Leu Val Leu His Asp		
180	185	190
Trp Gly Ser Ala Leu Gly Phe Asp Trp Ala Asn Gln His Arg Asp Arg		
195	200	205
Val Gln Gly Ile Ala Phe Met Glu Ala Ile Val Thr Pro Met Thr Trp		
210	215	220
Ala Asp Trp Pro Pro Ala Val Arg Gly Val Phe Gln Gly Phe Arg Ser		
225	230	235 240
Pro Gln Gly Glu Pro Met Ala Leu Glu His Asn Ile Phe Val Glu Arg		
245	250	255
Val Leu Pro Gly Ala Ile Leu Arg Gln Leu Ser Asp Glu Glu Met Asn		
260	265	270
His Tyr Arg Arg Pro Phe Val Asn Gly Gly Glu Asp Arg Arg Pro Thr		
275	280	285
Leu Ser Trp Pro Arg Asn Leu Pro Ile Asp Gly Glu Pro Ala Glu Val		
290	295	300
Val Ala Leu Val Asn Glu Tyr Arg Ser Trp Leu Glu Glu Thr Asp Met		
305	310	315 320
Pro Lys Leu Phe Ile Asn Ala Glu Pro Gly Ala Ile Ile Thr Gly Arg		
325	330	335
Ile Arg Asp Tyr Val Arg Ser Trp Pro Asn Gln Thr Glu Ile Thr Val		
340	345	350
Pro Gly Val His Phe Val Gln Gln Asn Ser Asp Gly Val Val Ser Trp		
355	360	365
Ala Gly Ala Arg Gln His Arg Arg Pro Gly Ser Ala Leu Ile Ser Arg		
370	375	380
Asp Gln Glu Cys Asp Phe Arg Arg Arg Arg Arg Pro Ala Cys Gln Leu		
385	390	395 400
Ile Arg Leu Pro Ala		

161

Gln Arg Ser Ser Tyr Leu Pro Ser Glu Leu Val Ala Ala Phe Leu Trp
 435 440 445

Ala Gln Phe Glu Glu Ala Glu Arg Ile Thr Arg Ile Arg Leu Asp Leu
 450 455 460

Trp Asn Arg Tyr His Glu Ser Phe Glu Ser Leu Glu Gln Arg Gly Leu
 465 470 475 480

Leu Arg Arg Pro Ile Ile Pro Gln Gly Cys Ser His Asn Ala His Met
 485 490 495

Tyr Tyr Val Leu Leu Ala Pro Ser Ala Asp Arg Glu Glu Val Leu Ala
 500 505 510

Arg Leu Thr Ser Glu Gly Ile Gly Ala Val Phe His Tyr Val Pro Leu
 515 520 525

His Asp Ser Pro Ala Gly Arg Arg
 530 535

2) INFORMATION FOR SEQ ID NO:162:

- (A) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 294 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

X1 SEQUENCE DESCRIPTION: SEQ ID NO:162:

Asn Glu Ser Ala Pro Arg Ser Pro Met Leu Pro Ser Ala Arg Pro Arg
 1 5 10 15

Tyr Asp Ala Ile Ala Val Leu Leu Asn Glu Met His Ala Gly His Cys
 20 25 30

Asp Phe Gly Leu Val Gly Arg Ala Trp Asp Ile Val Thr Asp Ala Ala
 35 40 45

His Asp Asp Arg Ala Gly Leu Gly Val Asp Glu Gln Phe Arg His Val
 50 55 60

Gly Phe Leu Glu Pro Ala Pro Val Leu Val Asp His Arg Asp Asp Leu
 65 70 75 80

Gly Gly Leu Thr Val Asp Trp Leu Val Ser Trp Pro Arg Gln Arg Gly
 85 90 95

Asp Lys Asp Val Val Leu Gln Arg His Trp Leu Ala Leu Arg Arg Ser
 130 135 140
 Glu Thr Leu Glu His Thr Pro His Gly Arg Arg Pro Val Arg Pro Arg
 145 150 155 160
 His Arg Gly Asp Asp Arg Phe His Glu Arg Asp Pro Leu His Ser Val
 165 170 175
 Ala Met Leu Val Ser Pro Val Glu Ala Glu Arg Arg Ala Pro Val Val
 180 185 190
 Gln His Gln Tyr His Val Val Ala Glu Val Glu Arg Ile Pro Glu Arg
 195 200 205
 Glu Gln Lys Val Ser Leu Leu Ala Ile Ala Ile Ala Val Gly Ser Arg
 210 215 220
 Trp Ala Glu Leu Val Arg Arg Ala His Pro Asp Gln Ile Ala Gly His
 225 230 235 240
 Gln Pro Ala Gln Pro Phe Gln Val Arg His Asp Val Ala Pro Gln Val
 245 250 255
 Arg Arg Arg Gly Val Ala Val Leu Lys Asp Asp Gly Val Thr Leu Ala
 260 265 270
 Phe Val Asp Ile Arg His Ala Leu Pro Gly Asp Phe
 275 280

1. INFORMATION FOR SEQ ID NO:163:

1. SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 264 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:163

ATGACATCTT CTCTCTCTCTT TTTTGGCAAG AGTTTTCCTT GATTCTCTCTT CTACTCTCTCT
 TCCATGCGAG CGATCTCTCTT TTTCTCTCTAT TCTTCTCTCTT AAGATCTCTT TCTTCTCTCTA
 ATCTCTCTCTT TGGTCTCTCTA TTTCTCTCTCTT AAGATCTCTT TCTTCTCTCTT TCTTCTCTCTA
 TCTCTCTCTA TCTCTCTCTCT TTTCTCTCTCTT TCTCTCTCTA TCTCTCTCTT TCTCTCTCTA
 TCTCTCTCTT TCTCTCTCTT TCTCTCTCTT

(B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:164:

TAGTCGGCGA CGATGACGTC GCGGTCCAGG CCGACCGCTT CAAGCACCAC CGCGACCACG	60
AAGCCGGTGC GATCCTTACC CGCGAACCAG TGGGTGAGCA CCGGGCGTCC GCGGGCAAGC	120
AGTGTGACGA CACGATGTAG CGCGCGCTGT GCTCCATTGC GCGTTGGGAA TTGGCGATAC	180
TGCTCGGTCA TGTAGCGGGT GCGCGCGTCA TTTATCGACT GGCTGGATTG CCGCGACTCG	240
CGGTTGGACC CGTCATTGGT TAGCAGCCTC TTGAATGCGG TTTCGTGCGG CGCTGACTCG	300
TGGCGCTCAT CATCGCGAG CTCGGGGAAC GCGAGCAGGT GGACGTGAT GCGCTCGGA	360
ACCGCTCCTG GACCGCGCGG GCGAAGCTCC CCGGACGACC GCAGGTGCGG AACGTGCGTG	420
ATCGCGACGC GCGGCACCGT TCGCGCTTCT GCGGAATTCC GCACGAGGCT GCGGAGCCAC	480
CGCGCATCAC CAAGCAACGC TTGCGCAGTA CGGATCGTCA CTTCCGCATC CGGCAGACCA	540
ATCTCGTCCG CCGCCATCCT CAGATCCCGC TCGTCCCTTG ACAAGAACGG CCGCAGATGT	600
CGCAGCGGCT ATCGGAGATT GAACCGGCA GCGAGTTCTT CAATCGCTGC GCGCTCCCGC	660
AGTATTGCA GTTTCGCGCG CTCGCGGTAT TTAGCAAGCA TCGGAGTCTC GACGAACTCG	720
CGCCACGTAA CGCACGGCGT AGCTCCCGCG CTGACGCGGA CGATCGGCGG GTGATCTTTG	780
CGCGAGCTCT CTAGCGGTT GATCCACCGC TTCTCGGTTC GCGCGCGGAG CGCGATCAGC	840
TTATCGACCT CGGCCTATGC CGACGGCAAG CTGCGCGGCT TCGTCGAGGT CAAGAATCTC	900
ACCATCGGCA CGGGCAGCAA GGTGCGCGAC CTGAGCTAGC TCGCGACGC CGACATCGGC	960
GATACAGCA ACATCGGCT TCGAGCTTC TTCTCAAT AGGAGGTAC CTGCAACCG	1020
CGGAGCTTC TCGGTTCGA CTAGCGAGT GGTTCGAGA CATTTCTT CGCGGACTA	1080
AGTATGCTG ACGGCGCTA TCGCGGCTT GCGAGCTTC TCGGAGGA TCGCGCGCG	1140
CGCGCGCTG CAGTGTGCGG GGTTCGGA	1200

1 INFORMATION FOR SEQ ID NO:164:

1. SEQUENCE CHARACTERISTICS

A. LENGTH: 1200 bases

GCAAAGGCGG CACGGGCGGG GCCGGCATGA ACAGCCTCGA CCGGCTGCTA GCGGCCCAAG 60
 ACGGCGGGCCA AGGCGGCACC GCGGCGACCG GCGGCAACGC CGGCGCGGGC GGCACCAAGT 120
 TCACCCAAGG CGCCGACGGC AACGCCGGCA ACGGCGGTGA CGGCGGGGTC GCGGCAACG 180
 GCGGAAACGG CGGAAACGGC GCAGACAACA CCACCACCGC GCGCGGCC 227

(2) INFORMATION FOR SEQ ID NO:166:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 304 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:166:

GCTGGCCACC ATGGGCGGGC AGGGCGGTAG CGGTGGCCGC GGCTCTACCC CAGGCGCCAA 60
 GGGCGCCCAAC GGCTTCACTC CAACCAGCGG CCGGCGACGGC GCGGACGGCG GCAACGGCGG 120
 CAACTCCCAA GTGGTCGGCG GCAACGGCGG CCGGCGGGCC AATGGCGGCA ACGGCGGCAG 180
 GCGGCGACCG GCGGCGAACC GCGGCGCGCG GCGGCGACGGC GCGTTTGGTG GCATGAGTGC 240
 CAACGGCACC AACGTTGGTG AAAACGGGCG AAACGGTAAC CCGGCGGGCA ACGGTGGCGC 300
 CGGC 304

(2) INFORMATION FOR SEQ ID NO:167

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1439 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:167

ATGGAAGGCT GCGGAGGCTC TATTAACAGG AATACATGCA CAGGAGGCTC ATGAGTGAAC 60
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 120
 TATGAGGCTC GGTCTAGGAA TATTTTTCG CAGGAGGCTC TCGGCGGAA GCGGAGGCGG 180
 GCGGAGGCTC GGTCTAGGAA TATTTTTCG CAGGAGGCTC TCGGCGGAA GCGGAGGCGG 240
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 300
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 360
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 420
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 480
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 540
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 600
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 660
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 720
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 780
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 840
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 900
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 960
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 1020
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 1080
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 1140
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 1200
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 1260
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 1320
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 1380
 TATGAGGCTC ATTTAACAAT GCGGCTTCA CCGGCGACGG CAGGAGGCTC GCGGAGGATC 1440

GGCTCGGGCG CCGATGGAGT GATACCTTCG CCGCGGACCA GCACCCGGAC GTGCAGATGG 480
 ACTACGTGAT GGCCAATCCG CCGTTCACCA TCAAAGACTG GCGCCGCAAC GAGGAAGACC 540
 CACGCTGGCG CTTCGGTGTG CCGCCCGCCA ATAACGCCAA CTACGCATGG ATTTCAGACA 600
 TCCTGTACAA CTTGGCGCCG GGAGGTCGGG CCGGCGTGGT GATGGCCAAC GGGTCGATGT 660
 CGTCCAACTC CAACGGCAAG GGGGATATTC GCGCGCAAAT CGTGGAGGCG GATTTGGTTT 720
 CCGTCATGGT CCGCTTACCC ACCCAGCTGT TCCGCAGCAC CCGAATCCCG GTGTCCCTGT 780
 GGTTTTTTGC CAAAAACAAG GCGGCAGGTA AGCAAGGGTC TATCAACCGG TCGGGCAGG 840
 TCGTGTTCAT CGACGCTCGT GAACTGGGCG ACCTAGTGGA CCGGGCCGAG CCGGCGCTGA 900
 CCAACGAGGA GATCGTCCGC ATCGGGGATA CTTTCCACGC GAGCAGGACC ACCGGCAACG 960
 CCGGCTCCCG TGGTCCCGCG GGTAAATGGG GCACTGGGCT CAACGGCGCG GCGGCTGCTC 1020
 CCGGGGCGCG CCGCAACGCG GGTGTGCGCG GCGTGTCTTT CCGCAACGCT GTGGGCGGCG 1080
 ACGGGCGCAA CCGCGGCAAC GCGGCGGACG GCGGCGACGG CACGACGGCG GCGGCGGCGG 1140
 GCAAGGCGCG CAACGGCAGC AGGGGTGCGG CAGCGGGCTC AGGCGTCGTC AACGTCACCG 1200
 CCGGTCACCG CCGCAACGCG GCGAATGCGG GCAACGGCGG CAACGGCTCC GCGGGCGCGG 1260
 CCGGCGAGGG CGGTCCCGCG GCGAGGCGCG GCAACGGCGG CACGCGCGCG GTGCGCACCG 1320
 CCGGCGCGAG CCGCAAGGCG GCGAACGGCA CAGCGGGTGG CCGGCGCGCG TCAGGCGCTC 1380
 TCAACGTCAG TCGGCGCGCG GCGGCGAAGG CCGGCAATCG CCGCAACGCG GCGAACGCG 1440

2. INFORMATION FOR SEQ ID NO.168:

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 322 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO.168

GCGCGCGCGG GCGCGGATTT TGTGTGCTCT TCAATGTCTG TCGGATACAG GCGCGTCAAT 40
 TCGGTAACCG CCGGATGGGT GCGGCTGCGG TCGGTGCGCG CCGGATGCGG GCGCGCGCGG 80
 TCGTCAGGCT TCGGTGCTCT TCAAGGCTCT TCGGATGCGG TCGGATGCGG TCGGATGCGG 120

(2) INFORMATION FOR SEQ ID NO:169:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 80 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:169:

GCAACGGTGG CAACGGGGGC ACCAGCACGA CCGTGGGGAT GGCGGAGGT AACTGTGGTG 60
 CCGCGGGGCT GATCGGCAAC 80

(2) INFORMATION FOR SEQ ID NO:170:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 392 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:170:

GGGTGTGTG GCACTCACAC CGCGGATTC GCGGACGTTG GCGCGCCAAT ATCCAGGTCA 60
 AGGCGTACTA GTTACCGTCC GAGGACCGCC GCATCAAGGT GCGGGTCAGC GCCCAAGGAA 120
 TCAAGGTGAT CGACCGCCAC GGGCATCGAG GCGGTGTTCC CCGCGGTCCG GCAGGATCCG 180
 CCGCGGGCCA GTTCGCGCCC CAGCGGGGCT CATGCTTCC AACCGCCCG ATCGTGTGAG 240
 TCAAGGTGAT GCGGCGCAC GAGATTCGTA TATTGTAA GCGGTGTTCC ACCGCAGGGA 300
 CCGGTATATC GATGTGTAAC GATGTGACT TCAAGAAC GGCATACCA TCCCGTGATC 360
 CCGCGACAGC CAGGAGTCC AAGACCGTTA CA 392

(2) INFORMATION FOR SEQ ID NO:171:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 535 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:171:

ATGCGTGA CCGGCGGA TTTTATG TTTTATG TTTTATG TTTTATG TTTTATG TTTTATG

GGCGGCACCG GCGGCACCGG CGGAGCGGGG GGAGCGGGCG GGGCGGGTGG CGCCATCGGT	240
ACCGGGCGCA CCGGCAGCGC GGTGGGCAGC GTGGTAACG CCGGGATCGG CGGTACCGGC	300
GGTACGGGTG GTGTGGGTGG TGCTGGTGGT GCAGGTGGG CTGCGGGCCG TGGCAGCAGC	360
GCTACCGGTG GCGCCGGGTT CGCGGGCGGC GCCGGCGGAG AAGGCGGACC GGGCGGCAAC	420
AGCGGTGTGG GCGGCACCAA CGGCTCCGGC GCGCGCGGCG GTGCAGGCGG CAAGGGCGGC	480
ACCGGAGGTG CCGCGGGGTC CGGCGCGGAC AACCGCACCG GTGCTGGTTT CCGCG	535

(2) INFORMATION FOR SEQ ID NO:172:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 590 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:172:

CGGAGGTGCG GGGGGCGATA GGGGGGTGAC CGACTACTAC ATCATCCGCA CCGAGAATCG	60
GCGGTGCTG CAACGGGTGC GGGGGGTGCC GGTTCATCGGA GATCCGCTGG CCGACCTGAT	120
CGAGCGGAAC CTGAAGGTGA TCGTCAATCT GGGGTACGGC GACCGGAATC ACGGCTACTC	180
GACGAGCTAC GCGGATGTGC GAAGGGCTTT TGGGCTGTGG CCGAACGTGC CGCCTCAGGT	240
GATCGCCGAT GCGGTGGCGG CGGGAACACA AGAAGGCTTC GTTGACTTCA CGGCGGACCT	300
TGAGGGGCTG TCGGGCGAAC GGTTCAGGCT TGTGAGATG GAGCTGGCGC AACCGCGCGA	360
TGTGGTGGCG GCGGTGGCGG CGGCAAGGAC TCGGGCGGAG GTGGTGAACA CGCTCGCCAG	420
GATCATCTCA AGCAACTACG CGTCTGTGTT GTGACCGTGC GACATCGCCC TCGCCTGGTC	480
AGCAAGCTGC GGTGTACAC AGCGCAATC TGTGTCAAGT AACTGCTTGC CGGCAATCTC	540
ATCAAAGCA TCGGTATCTC ATGCGGCTC AATGAGTTC CAATCAATAT GATAGCGCG	600
TGTTGTGAA TTGCTGACGC TCGTGGCGGC GGTGTGAA ATATTCGAA AGATGAGCG	660
TGTGTGACG TACCGGATTC CGGACGGCAT	720

(2) INFORMATION FOR SEQ ID NO:173:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 177 base pairs

```

ACGGTGACGG CGGTACTGGC GCGGGGACG GCGGCAACGG CGGGAATCCC GGGTGGCTCT      60
TGGGCACAGC CGGGGGTGGC GGCAACGGTG GCGCCGGCAG CACCGTACT GCAGGTGGCG      120
GCTCTGGGGG CACCGGCGGC GACGGCGGGA CCGGCGGGCG TGGCGGCCCTG TTAATGGGCG      180
CCGGCGCCCG CGGGCACGGT GGCCTGGCG GCGCGGGCGG TGCCGGTGTG GACGGTGGCG      240
GCGCCGGCGG GGCCGGCGGG GCGGGCGGCA ACGGCGGGCG CGGGGGTCAA GCCGCCCTGC      300
TGTTCCGGCG CGGCGGCACC GCGGGAGCCG GCGGCTACGG CGCGGATGGC GGTGGCGGCG      360
GTGACGGCTT CGACGGCAGC ATGCGCGGCC TGGGTGGTAC CGGTGGC      407

```

(2) INFORMATION FOR SEQ ID NO:174:

1. SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 468 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:174:

```

GATCGGTGAG CGCATGCGCC TCGCGCGCAA GCGATTCCGC GGTCTCACC GAGAACATCG      60
TGCACGCGGG GGCGCGGAC AGCCCGCTGC GGTGCGGGCG CTCGAACGCC TCCAGCAGGC      120
ACAGCCAGTC GTTGCGGGG TCGGAGGCGA ACACGTGCGT CTCACCGGTG TAGATGCGCG      180
GGATGCGCGC CTCGCGAAC GATTTCGGG ACGGCGGGCG GTCTTTGTGA TGTCTGACGA      240
TACGCGCGAT GTCTGCGCG AGCAGCGCG GCGCGCGGAA GTTGGCGCGG CTGCGGAGTA      300
GCGCGCGGAC CTCGCGGGG AGGTCTTGG GATGTGCGG CGGAGCGGT CCGCGCGGAC      360
GCGCGAAAAA CGACCGCTGA CGCATCTGG TCGGCTGGG ATATCGCTTG CGTCTCTGGG      420
GATATTGGA CGCGATGCG CGCATCTGG TCGGCTGGG ATATCGCTTG CGTCTCTGGG      468

```

(2) INFORMATION FOR SEQ ID NO:175:

1. SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 219 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:175:

GGCGGCGCGG GCGGCAATGG CGGCGCGGGC GGCAAGCGC AGGCGGCGCG GTACACCGAC 180
 GGCGGCGCGG GCACCGCGCG CGACGGCGGC AACGGCGGC 219

(2) INFORMATION FOR SEQ ID NO:176:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 494 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:176:

TAGCTCGCGG GAGGCGCGCA AGGCGCGCGA CGGTGGCCAC GGCGGTGACG GCGTCGGCGG 50
 TAACAGTTCC CTCACCAAG GCGGCAGCGG CGGTGGCGCG GCGCGCGCGG GCGCGCGCGG 120
 CAGCGGCTTT TTGCGCGCGA AGGCGCGGCTT CGCGCGCGAC GCGCGTCAGG GCGGCCCCAA 180
 GCGCGCGCGT ACCGTGCGCA CGGTGCGCGG TGGCGCGCGG AACGCGCGGTG TCGCGCGCGG 240
 GCGCGCGCGAC GCGCTCTTTG CGGTGCGCGG CGGCGAGCGG GCGCTCGGTG GCGAGGCGCG 300
 CAATGCGCGG GCGTCCACCG GCGGCAACCG CGGCTTTGGG GCGCGCGCGG GTGGCGGAGG 360
 CAACGCGCGG GCGCTCGCGG AATCGCGCGT GAGGTGCGAC AGCGCGCGCA AGTTGCGCTG 420
 CATCGATCA GCGCGCTACT GCGCGGAAAC GCGCGAACAT CAGCGGAGTT AGCGGCGCGG 480
 ATTTCTGAT CACG 494

(2) INFORMATION FOR SEQ ID NO:177:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 220 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:177:

CGGCGCGCGG GCGGCAATGG CGGCGCGGGC GGCAAGCGC AGGCGGCGCG GTACACCGAC 50
 GGCGGCGCGG GCACCGCGCG CGACGGCGGC AACGGCGGC 120
 GCGGCGCGG GCGGCAATGG CGGCGCGGGC GGCAAGCGC AGGCGGCGCG GTACACCGAC 180
 GGCGGCGCGG GCACCGCGCG CGACGGCGGC AACGGCGGC 220

- (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:178:

ATGGCGGCAA CGGGGGCCCC GGCGGTGCTG GCGGGGCGCG CGACTACAAT TTCCAACGGC	60
GGGCAGGGTG GTGCCGCGCG CCAAGGCGGC CAAGGCGGCG TGGGCGGGCG AAGCACCACC	120
TGATCGGCCT AGCCGCACCC GGGAAAGCCG ATCCAACAGG CGACGATGCC GCCTTCCTTG	180
CCGCGTTGGA CCAGGCGCGC ATCAGCTAGC CTGACCCAGG CCACGCCATA ACGGCCGCCA	240
AGGCGATGTG TGGGCTGTGT GCTAACGGCG TAACAGGTCT ACAGCTGGTC GCGGACCTGC	300
GGGACTACAA TCGGGGGTGT ACCCTGGACA GCGCGGCCAA GTTCGCTGCC ATCGCATCAG	360
GCGGCTACTG CCGCGAACAC CTGGAACA	388

(X2) INFORMATION FOR SEQ ID NO:179:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 400 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:179:

GCAAGGGCGG CACCGCGCGG GCGGGCATGA ACAGCCTCGA CCGGCTGCTA GCGGCCCAAG	60
ACGGGGGCCA AGGCGGCACC GCGGGCAGCG GCGGCACCGC CCGCGCCCGC GGCACCAAGT	120
TTAGGCAAGG GCGGCACCGC AACCGCGCGA AGGGCGGTGA GCGCGGGGTC GCGGCCAAGC	180
GCGAAACGGG GGGAAACGGC GCAGACAACA GAGCAACGGC GCGCGCCCGC ACCACAGGCG	240
TGAGGGGGG GCGCGCGCGG GCGGCGGAA GCGGGGAGG GCGGGCAGCG	300
TAGCGGGG GCGGCACCGC AACCGCGCGA AGGGCGGTGA GCGCGGGGTC GCGGCCAAGC	360
GCGAAACGGG GGGAAACGGC GCAGACAACA GAGCAACGGC GCGCGCCCGC ACCACAGGCG	420

(X2) INFORMATION FOR SEQ ID NO:180:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 598 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

AGCGGCAACG GCGGGCAACG GCGGCAGCGG GGGCAACGGC GGCAACGCGG GCATGGGCGG 120
 CAACAGCGGG ACCGGCAGCG GCGACGGCGG TGCGGGCGGG AACGGCGGGG CGGCGGGCAC 180
 GGGCGGCACC GCGGGCGAGG GCGGCCTCAC CGGTACTGGC GGCACCGGGG GCAGCGGTGG 240
 CACCGGCGGT GACGGCGGTA ACGGCGGCAA CGGAGCAGAT AACACCGCAA ACATGACTGC 300
 GCAGJCGGGC GGTGACGJTG GCAACGGCGG CGACGGTGCC TTCGGCGGGG GGGCCGGGGG 360
 CGGCGGCGGT GGCTTGACCG CTGGCGCCAA CGGCACCGGC GGGCAAGGCG GCGCCGGCGG 420
 CGATGGCGGC AACGGGGGCA TCGCGGGCCA CGGGGCACTC ACTGACGACC CCGGCGGCAA 480
 CCGGGGCACC GCGGCAACG GCGGCACCGG CGGCACCGGC GCGCGGGCA TCGGCAGC 538

2. INFORMATION FOR SEQ ID NO:181:

1. SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 339 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:181:

GGGCTGGTGG TGCGCGGGG CAGCTCTTCA GCGCGGAGG GCGGCGGGT GCGGTGGGG 60
 TTGGGGGAC GCGGGGCGAG GTGGGGGCTG GCGGTGCGCG AGCGGCGGGG GCGGACGCGG 120
 CCGCGAGGAT AGGTCTTACC GTGGGTACCG GTTGGGCTGG GCGGGCGGGG GCGGTGGGG 180
 GCGGGGCGG GAACGGGATT GCGGGCGGCA TGAACGGGCT GCGTGGTGGG GCGGCGAT 240

2. INFORMATION FOR SEQ ID NO:182

1. SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 985 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO:182

AGCAJCGCTA GCGGTGGGCT GGGGTGGGCT GCGGGGCTG GCGGAGAAAT GCGAGCGGGT 6
 GCGGAGAGCG GTGTGGGCTG GCGGAGAGCT TCGGGGCTG GCGGAGAGCT AGCGGCGAAG 12
 GCGGAGAGCG GCGGTGGGCT GGGGTGGGCT GCGGGGCTG GCGGAGAAAT GCGAGCGGGT 18

GACGGCCGGCC AAGGCGGGCCA AGGCGGGGGCC GGGCGGAGCG CCGCGGCCCGG CGGCATCAAC 420
 GGGGCGGGCG GGGCGGGCGG CAACCGCGGC GACGGCGGGG ACGGCGCAAC CGGTGCCGCA 480
 GGTCTCGGCG ACAACGGCGG GTCGGCGGT GACGGTGGGG CCGGTGGCGC CGCCGGCAAC 540
 GGGCGCAACG CGGGCGTCGG CTTGACAGCC AAGGCGGGCG ACGGCGGCGG CGCGGGCAAT 600
 GGGCGCAACG GGGGCGGGCG CGGTGCTGGC GGGGCGGGCG ACAACAATTT CAACGGCGGC 660
 CAGGGTGGTG CCGGCGGGCA AGGCGGGCAA GCGGGCTTGG GCGGGGCAAG CACCACCTGA 720
 TCGGCTTAGC CGCACCCTGG AAAGCGGATC CAACAGGGCA CGATGCGGCG TTCCTTGCGG 780
 CGTTGGACCA GGGCGGATC ACCTAGGCTG ACCGCGGCA CGCATAACG GCGGCGAAGG 840
 CGATGTGTGG GCTGTGTGT AAAGCGGCTA CAGGTCTACA GGTGGTGGCG GACCTGCGGG 900
 AATACAATTC CCGGCTGACC ATGGACAGCG CGGCGAAGTT CGCTGCCATC GCATCAGGCG 960
 CGTACTGCGC CGAACACCTG GAACA 985

2. INFORMATION FOR SEQ ID NO:183:

(A) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2138 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(B) SEQUENCE DESCRIPTION: SEQ ID NO:183:

CGGCACGAGG ATGGGTACCG CGGCGGATCG CGAGTTCGCG ATTGCGCGGGG TTTGCGGACG 50
 CGAGGAAAGC CGGTACGAGA TGGCGGTGCG GAAGTAGGCG GATCGGTTGCG CGATGCGGCG 100
 ATGACCGGCG CGCATCAAT TGTGCGGCA ACCTTTGAGT TTAGCGACGA TAATGGCTAT 150
 TTTACTTAGT AGGATGATCG GATATGAGCG AATGCGCA TGTACCGGTG GATCAGCAAG 200
 AATGTTGAA GAGGCGGAA TATGTTGAA TGTGATCG TACGCGGCGG AATGATGTCG 250
 TGTGCGGAGC GTGCGAATG AATGCGGCTA AAAAGCGGT TCGACAGCTG GTATTGTGCG 300
 TCGACGAGAT CGGCGAATAG TGTGCGGCG GTGCGAAGA GCGCGAGCGT CTGCGGACCT 350
 TGTGCGGAA CGGCGGCAAG TGTGATGCG ATGTTGATGA GAGGCTTGG ACCGCGCTCG 400
 AATGATGAGG CGAAGCAAT TGTGCGGCG ATGTTGATGA TGTGATGAGG TGTGATGAGG 450

CGGATGGGTG GAACACTTTC AACCTGACGC TGCAAGGCGA CGTCAAGCGG TTCCGGGGGT 720
 TTGACAACTG GGAAGGCGAT GCGGCTACCG GTTGGGAGGC TTGCTCGAT CAACAACGGC 780
 AATGGATACT CCACATGGCC AAATTGAGCG CTGGGATGGC CAAGCAGGCT CAATATGTGG 840
 CGCAGCTGCA CGTGTGGGCT AGGCGGGAAC ATCCGACTTA TGAAGACATA GTCCGGGCTCG 900
 AACGGCTTTA CGCGGAAAAC CCTTCGGCCC GCGACCAAAT TCTCCCGGTG TACGCGGAGT 960
 ATCAGCAGAG GTCCGACAAG GTGCTGACCG AATACAACAA CAAGGCAGCC CTGGAACCGG 1020
 TAAACCCGCC GAAGCCTCCC CCGGCCATCA AGATCGACCC GCGCCCGCCT CCGCAAGAGC 1080
 AGGGATTGAT CCTGGCTTC GTGATGCCGC GTTCTGACCG CTCCGGTGTG ACTCCCGGTA 1140
 CCGGGATGCC AGCCGCACCG ATGGTTCCGC GTACCGGATC GCGGGGTGGT GGCCTCCCGG 1200
 CTGACACGGC GCGCGAGCTG ACCTCGGCTG GCGCGGAAGC CGCAGCGCTG TCGGGCGACG 1260
 TCGCGGTCAA AGCGGCATCG CTCCGTGGCG GTGGAGGCGG CCGGGTCCCG TCGGCGCCGT 1320
 TCGGATCCCG GATCGGGGGC GCGAATCCG TCGGGCCCGC TGGCGCTGGT GACATTGCCG 1380
 GCTTAGGCCA GGAAGGGCC GCGCGCCCGC CCGCGCTGGG CCGCGGTGGC ATGGGAATGC 1440
 GATGGGTGC CCGGCATCG GAGCAAGGGC CCGCCAAGTC CAAGGGTTCT CAGCAGGAAG 1500
 AGGAGCGCT CTACACCGAG GATCGGGCAT GCACCGAGG CTGATTGGT AACCGTCGGC 1560
 GGAAGACG TAAGGAGTGG AAGTGAGCAT GCACGAATG GACCGGATG TCGCCCGGGC 1620
 GTTGAAGTC CCGCGCGCTT TTGAGTGGC GTAGACGGG ACCTCAATC AGATGACAA 1680
 TGGATGTTG CCGGCCACCG ACGAAGCGA GACCTCGAA GTGACGATCA ATGGGCACCA 1740
 GTGGTACG GCGCTCGCA TCGAAGATG TTTGCTGAAG AAGCTGGGTG CCGAGGCGGT 1800
 GTTGAAGG GTCAACGAG GTGTGACAA TTTGAGGCT CCGGCTGCT GTTATAACGA 1860
 GTTGTGCT GAGGAGCTGA GTGTGCTT ATGAGGATG TCGGCGGCA TGAAGGAGG 1920
 AATGGCTAA GTTATTGCT GTGTGCTA TATGACGA CCGAATGAG GTTCAATGC 1980
 GTTATGAA GTGCGCGAG AGCGCTGA TATGATGCT CAATGTTTT AGATGGATG 2040
 CCGGGTTCG GAGGCGGCA TATGCTGCT GTTATATT CCGGCAGCTA GTGTGCTTA 2100
 GTTGTGTA GTTGTGTA GTATGCTT GTTATGA 2160

(C) STRANDEDNESS:

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:184:

Met Thr Gln Ser Gln Thr Val Thr Val Asp Gln Gln Glu Ile Leu Asn
 1 5 10 15
 Arg Ala Asn Glu Val Glu Ala Pro Met Ala Asp Pro Pro Thr Asp Val
 20 25 30
 Pro Ile Thr Pro Cys Glu Leu Thr Ala Ala Lys Asn Ala Ala Gln Gln
 35 40 45
 Leu Val Leu Ser Ala Asp Asn Met Arg Glu Tyr Leu Ala Ala Gly Ala
 50 55 60
 Lys Glu Arg Gln Arg Leu Ala Thr Ser Leu Arg Asn Ala Ala Lys Ala
 65 70 75 80
 Tyr Gly Glu Val Asp Glu Glu Ala Ala Thr Ala Leu Asp Asn Asp Gly
 85 90 95
 Glu Gly Thr Val Gln Ala Glu Ser Ala Gly Ala Val Gly Gly Asp Ser
 100 105 110
 Ser Ala Glu Leu Thr Asp Thr Pro Arg Val Ala Thr Ala Gly Glu Pro
 115 120 125
 Asn Phe Met Asp Leu Lys Glu Ala Ala Arg Lys Leu Glu Thr Gly Asp
 130 135 140
 Gln Gly Ala Ser Leu Ala His Phe Ala Asp Gly Trp Asn Thr Phe Asn
 145 150 155 160
 Leu Thr Leu Gln Gly Asp Val Lys Arg Phe Arg Gly Phe Asp Asn Trp
 165 170 175
 Gln Gly Asp Ala Ala Thr Ala Val Ala Ala Ser Leu Asp Gln Gln Arg
 180 185 190
 Gln Thr Gln Leu His Val Ala Lys Leu Ser Ala Ala Met Ala Lys Gln
 195 200 205
 Ala Gln Tyr Val Ala Gln Leu His Val Trp Ala Arg Arg Gln His Pro
 210 215 220
 Thr Tyr Glu Asp Ile Val Gly Leu Gln Arg Leu Tyr Ala Gln Asn Pro
 225 230 235 240

175

Val Asn Pro Pro Lys Pro Pro Pro Ala Ile Lys Ile Asp Pro Pro Pro
 275 280 285
 Pro Pro Gln Glu Gln Gly Leu Ile Pro Gly Phe Leu Met Pro Pro Ser
 290 295 300
 Asp Gly Ser Gly Val Thr Pro Gly Thr Gly Met Pro Ala Ala Pro Met
 305 310 315 320
 Val Pro Pro Thr Gly Ser Pro Gly Gly Gly Leu Pro Ala Asp Thr Ala
 325 330 335
 Ala Gln Leu Thr Ser Ala Gly Arg Glu Ala Ala Ala Leu Ser Gly Asp
 340 345 350
 Val Ala Val Lys Ala Ala Ser Leu Gly Gly Gly Gly Gly Gly Val
 355 360 365
 Pro Ser Ala Pro Leu Gly Ser Ala Ile Gly Gly Ala Glu Ser Val Arg
 370 375 380
 Pro Ala Gly Ala Gly Asp Ile Ala Gly Leu Gly Gln Gly Arg Ala Gly
 385 390 395 400
 Gly Gly Ala Ala Leu Gly Gly Gly Gly Met Gly Met Pro Met Gly Ala
 405 410 415
 Ala His Gln Gly Gln Gly Gly Ala Lys Ser Lys Gly Ser Gln Gln Glu
 420 425 430
 Asp Glu Ala Leu Tyr Thr Glu Asp Arg Ala Trp Thr Glu Ala Val Ile
 435 440 445
 Gly Asn Arg Arg Arg Gln Asp Ser Lys Glu Ser Lys
 450 455 460

7 INFORMATION FOR SEQ ID NO:185

- 1. SEQUENCE CHARACTERISTICS
- A. LENGTH: 460 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS:
- D. TOPOLOGY: linear

81. SEQUENCE DESCRIPTION: SEQ ID NO:185

Ala Gly Asn Val Thr Ser Ala Ser Gly Ser Met Arg Phe Gly Ala Pi
 1 10 15

176

Phe Pro Pro Ser Arg Arg Gln Leu Arg Arg Arg Val Ser Arg Glu Ala
 50 55 60
 Thr Thr Arg Arg Ser Gly Arg Arg Asn His Arg Cys Gly Trp His Pro
 65 70 75 80
 Gly Thr Gly Ser His Thr Gly Ala Val Arg Arg Arg His Gln Glu Ala
 85 90 95
 Arg Asp Gln Ser Leu Leu Leu Arg Arg Arg Gly Arg Val Asp Leu Asp
 100 105 110
 Gly Gly Gly Arg Leu Arg Arg Val Tyr Arg Phe Gln Gly Cys Leu Val
 115 120 125
 Val Val Phe Gly Gln His Leu Leu Arg Pro Leu Leu Ile Leu Arg Val
 130 135 140
 His Arg Glu Asn Leu Val Ala Gly Arg Arg Val Phe Arg Val Lys Pro
 145 150 155 160
 Phe Glu Pro Asp Tyr Val Phe Ile Ser Arg Met Phe Pro Pro Ser Pro
 165 170 175
 His Val Gln Leu Arg Asp Ile Leu Ser Leu Leu Gly His Arg Ser Ala
 180 185 190
 Gln Phe Gly His Val Glu Tyr Pro Leu Pro Leu Leu Ile Glu Arg Ser
 195 200 205
 Leu Ala Ser Gly Ser Arg Ile Ala Phe Pro Val Val Lys Pro Pro Glu
 210 215 220
 Pro Leu Asp Val Ala Leu Gln Arg Gln Val Glu Ser Val Pro Pro Ile
 225 230 235 240
 Arg Lys Val Arg Gln Arg Lys Ala Leu Val Ala Arg Phe Gln Leu Pro
 245 250 255
 His Arg Phe Phe Gln Thr His His Val His Phe Thr His Arg His Val
 260 265 270
 His Arg Arg Ile Gly
 275

3. INFORMATION FOR SEQ ID NO:186:

SEQUENCE CHARACTERISTICS

A. LENGTH: 100 amino acids

177

Arg Val Ala Ala Ser Phe Ile Asp Trp Leu Asp Ser Pro Asp Ser Pro
 1 5 10 15
 Leu Asp Pro Ser Leu Val Ser Ser Leu Leu Asn Ala Val Ser Cys Gly
 20 25 30
 Ala Glu Ser Ser Ala Ser Ser Ser Ala Arg Ser Gly Asn Gly Ser Arg
 35 40 45
 Trp Thr Ser Met Pro Ser Gly Thr Arg Pro Gly Pro Arg Arg Ala Thr
 50 55 60
 Ser Arg Asp Asp Arg Arg Ser Ala Thr Ser Val Ile Pro Ser Arg Arg
 65 70 75 80
 Ser Val Ala Pro Arg Ala Glu Phe Gly Thr Arg Leu Ala Ser His Arg
 85 90 95
 Ala Ser Pro Ser Asn Ala Cys Pro Val Arg Ile Val Thr Ser Ala Ser
 100 105 110
 Gly Arg Pro Ile Ser Ser Pro Pro Ile Val Arg Ser Arg Ser Cys Val
 115 120 125
 Asp Lys Asn Gly Arg Arg Cys Ala Ser Gly Tyr Arg Arg Leu Asn Arg
 130 135 140
 Ala Arg Ser Ser Ser Ile Ala Ala Arg Cys Arg Thr Ile Gly Thr Phe
 145 150 155 160
 Arg Arg Ser Arg Tyr Ser Ala Ser Met Arg Val Ser Thr Asn Ser Pro
 165 170 175
 His Val Thr His Gly Val Ala Pro Gly Val Thr Arg Arg Ile Gly Gly
 180 185 190

INFORMATION FOR SEQ ID NO 137

1. SEQUENCE CHARACTERISTICS

- (A) LENGTH: 196 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

2. SEQUENCE DESCRIPTION: SEQ ID NO 137

His Ala Arg Asp Gly Met Asp Ile Arg Val Ser Gly Ile Gly Pro Arg
 1

178

35	40	45
Arg Asn Pro Arg Arg Ser Ser Arg Arg Asp Ala Glu Asp Arg Arg Val 50 55 60		
Ile Phe Ala Ala Thr Leu Val Ala Val Asp Pro Pro Leu Arg Gly Ala 65 70 75 80		
Gly Gly Glu Ala Asp Gln Leu Ile Asp Leu Gly Val Cys Arg Arg Gln 85 90 95		
Ala Gly Arg Val Arg Arg Gly Gln Glu Leu His His Arg His Arg His 100 105 110		
Gln Gly Ala Ala Pro Asp Leu Arg Arg Arg Arg Arg His Arg Arg Val 115 120 125		
Gln Gln His Arg Arg Leu Gln Arg Val Arg Gln Leu Arg Arg Tyr Val 130 135 140		
Gln Thr Ala His His Arg Arg Phe Ala Arg Thr Asp Arg Val Arg His 145 150 155 160		
His Val Arg Gly Pro Ser Asn His Arg Arg Arg Arg Val Tyr Arg Gly 165 170 175		
Arg His Ser Gly Ala Gly Gly Cys Pro Ala Gly Gly Ala Gly Ser Val 180 185 190		
Gly Gly Ser Ala 195		

INFORMATION FOR SERIAL NO. 134

2. SEQUENCE CHARACTERISTICS

- (A) LENGTH: 311 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

SEQUENCE SECTORIALITY

Val Arg Asp His Thr Seru Ala Leu Val Leu Met Val Glu Phe Leu Thr
5 10 15
Ser Thr Asn Ala Pro Ser Leu Ile Ser Ala Tyr Ala Glu Val Asp Lys
20 25 30
Leu His Gly Leu Leu Val Asp Thr Ala Leu Arg Thr Ile Asn Glu Thr
35 40 45

65 70 75 80

Ala Glu Tyr Arg Asp Arg Arg Lys Val Pro Ile Val Arg Gln Arg Ala
85 90 95

Ala Ile Glu Glu Leu Arg Ala Arg Phe Asn Leu Arg Tyr Pro Leu Ala
100 105 110

His Leu Arg Pro Phe Leu Ser Thr His Glu Arg Asp Leu Thr Met Gly
115 120 125

Gly Glu Glu Ile Gly Leu Pro Asp Ala Glu Val Thr Ile Arg Thr Gly
130 135 140

Gln Ala Leu Leu Gly Asp Ala Arg Trp Leu Ala Ser Leu Val Pro Asn
145 150 155 160

Ser Ala Arg Gly Ala Thr Leu Arg Arg Leu Gly Ile Thr Asp Val Ala
165 170 175

Asp Leu Arg Ser Ser Arg Glu Val Ala Arg Arg Gly Pro Gly Arg Val
180 185 190

Pro Asp Gly Ile Asp Val His Leu Leu Pro Phe Pro Asp Leu Ala Asp
195 200 205

Asp Asp Ala Asp Asp Ser Ala Pro His Glu Thr Ala Phe Lys Arg Leu
210 215 220

Leu Thr Asn Asp Gly Ser Asn Gly Glu Ser Gly Glu Ser Ser Gln Ser
225 230 235 240

Ile Asn Asp Ala Ala Thr Arg Tyr Met Thr Asp Glu Tyr Arg Gln Phe
245 250 255

Pro Thr Arg Asn Gly Ala Gln Arg Ala Leu His Arg Val Val Thr Leu
260 265 270

Leu Ala Ala Gly Arg Ile Val Leu Thr His Tyr Phe Ala Gly Lys Asp
275 280 285

Arg Thr Gly Phe Val Val Ala Leu Val Leu Val Ala Val Gly Leu Asn
290 295 300

Arg Asp Val Ile Val Ala Asp
305 310

2 INFORMATION FOR SEC ID NO 184

Journal of Management Studies, 19(1), 67-80.

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:189:

CTCGTGCCGA TTCGGCACGA GGTGAGCAGC CCAAGGGGCC GTTCGGCGAA GTCATCGAGG	60
CATTGCGCGA CGGGCTGGCC GGCAAGGGTA AGCAAATCAA CACCACGCTG AACAGCCTGT	120
CGCAGGCGTT GAACGCCTTG AATGAGGGCC GCGGCGACTT CTTCGCGGTG GTACCGAGCC	180
TGGCGCTATT CGTCAACGCG CTACATCAGG ACGACCAACA GTTCGTCCGG TTGAACAAGA	240
ACCTTGCGGA GTTCACCGAC AGGTTGACCC ACTCCGATGC GGACCTGTGG AACGCCATCC	300
AGCAATTGCA CAGCTTGCTC GCGGTGCGGC GCGGCTTCTT CGCCAAGAAC CGCGAGGTGC	360
TGACCGATGA CGTCAATAAT CTCGGGACCG TGACCACCAC GTTGCTGCAG CCGGATCCGT	420
TGGATCCCTT CGAGACGCTC TTGACATCTT TCCCGACGCT GGCGGCGAAC ATTAACCAGC	480
TTTACCATCC GACACACGGT GCGGTGGTGT CCGTTTCCGC GTTCACGAAT TTGCGCAACC	540
CAATGAGTT CATCTGCAGC TGATTCAGG CCGGTAGCTG GCTCGGTTAT CAAGAGTCGG	600
CCGAATCTG TCCGAGTAT CTGGCGCCAG TCCTCGATGC GATCAAGTTC AACTACTTTC	660
CGTTGGGCTT GAACGTGGCC AGCAGCGCTT CGACACTGCC TAAAGAGATC GCGTACTCCG	720
AGCGCGGCTT CGAGCGCGCC AACGGGTACA AGGACACCAC GGTGCGCGGC ATCTGGGTGC	780
CGGATACGCT GTTGTACAC CGCAACACCG AGCGCGGTTG GGTGGTGGCA CCGGGGATGC	840
AGGGGCTTCA GGTGGGACCG ATCAGCCAGG GTTGTGTGAC CCGCGAGTCC CTGGCGGAAC	900
TCATCGGTTG TCGGATATC CGCGCTCGCT CCGTGGGCTT GGAACCGCG CCGGACGCT	960
CGAATGCTTA CGACGAGTAC CCGGTGCTGC CGCGATCGG TTTACAGGCC CCACAGGTGC	1020
CGATACACCG CGCGCTGCTT GCGCGCGACG TAATCGCGCG TCGGTGCCA CCGGTCTTGG	1080
CGCGATGCTT GTTGGGAAGA GATCGCGCG GAGCGTGGG AAGTTGGAC TACATGGGCT	1140
CTTGTGCTCT TCGCGGCTT GTGCGACCT TCGGTGCTT TCGGTGCTT AGCGCGCGCT	1200
CTGGAAGGAT CGCGATGCG GAGTTGCTTA TACGCGGAT TACGCGGCT GCGTTGATGC	1260
CGCGATGCTT CGCGATGCG TCGTACGCGA CAGAACATGC TCGTACGAG ATCGCGCTTGT	1320
CGCGAAGCG AGCGGTGCG CAGCGGACGA TGACGATGAT GGTGCTGCTC CTGCGGCTGT	1380
CGCGGCTTCT GTGCTGCTT CGAGATGCTT CGAGGAAAT GTTGAAGAA TCGCGATGCT	1440
AACTGCGCTT CGATGCTTCT TCGCGGCTT TCGCGGCTT TCGCGGCTT TCGCGGCTT	1500

TGGCGACCCG GCTGGCAATC ATGGGUAATG GCATGGGCTG CTCCATGATG CCACTGTCCG 1680
 GGGCGGCACT GCAGACCCCTG GGGCCACATC AGATCGCTCG CGGTTCGACG CTGATCAGCG 1740
 TCAACCAGCA GGTGGGCGGT TCGATAGGGA CCGCACTGAT GTCGGTGCTG CTCACCTACC 1800
 AGTTCAATCA CAGCGAAATC ATCGCTACTG CAAAGAAACT CGCACTGACC CCAGAGAGTG 1860
 GCGCCGGGGG GGGGGCGGGG GTTGACCCCTT CCTCGCTACC GCGCCAAACC AACTTCGCGG 1920
 CCCAACTGCT GCATGACCTT TGGCACGCTT ACCCGGTGGT ATTCTGTATA GCGACCGCGC 1980
 TAGTGGTCTG GACGCTGATC CCGCGCGCAT TCCTGCCGAA ACAGCAGGCT AGTCATCGAA 2040
 GAGCACCGTT GGTATCCCA TACCTCTCTC TT 2072

2. INFORMATION FOR SEQ ID NO:190:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1923 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:190:

TCACCCCGCA GAAGTCGTTT GTGACGACG TGGACATCGA CTGGCTGTCTG ATGGTCGAGA 60
 TGGCGGTGCA GACCGAGGAC AACTAGGCGG TCAAGATCCC TGACGAGGAC CTCGCGGGTC 120
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 180
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 240
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 300
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 360
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 420
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 480
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 540
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 600
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 660
 TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT TGGTACGCTT 720

GACCGCGCGC AGGTCCGCGAG TCAGCAGCTT TCCGCGCGCA GCTTTCCGCG TGAAGCCGAC 900
 CAGGGCATCG TAGGTTCCGC CACCGGTGAC ATCGTGCTCG GCGAGGTGGT CCGTCAAGCC 960
 GCGATATGAG CAGGCATCCA GTGCCAGGTA GTTGCTGSAG GTGATGTCCG CCAAGTAGGC 1020
 GTGGACGGCA ACAGGGGCAA TACGATGCGG CCGTGGTAGC CCGGTCAAGA CCGAATAGGT 1080
 TTCCACAGCC GCGTGCGCGA TCAGATGGAC GCCACGGTTG AGCGCGCGCA CGGCGGCCTC 1140
 GTGCCCTTCG TGCCAGGTG CGAATCCGGC AACCAGCAGC CTGGTGTCTG GTGCGATCAC 1200
 CCGCGTGTGC GATCGAGCCT TTCCCGAAGC ATTTCTGTCG TCAACGGGGG CAGGGGACGT 1260
 TCTGCGCGTG CGACGAGAAG CGAGCCTTCC CGAACCAGTT CGACACCGGT CGGGGCGGGC 1320
 TCAATCTCGA TCCGCGCATC GCGCTCGGTG ATCTCCACCT GGTGCTTCCC GCGCAAGCCA 1380
 AGGCGCTCCG GAATCCGCTT GGAATCACC AGACGCTCTG CGACATCGAT GCTTGTTCGG 1440
 ATGCTAGGAA ATTTAGCATC GCACCTTCCA TAGGCGTCTC CTGCGCGGGA TGTGGGAGC 1500
 ATCGGTAGC GTATCGAAGC ATGTTTTCGG AAATGGCTGA GGGAGCGTGC GGTGCGCGTG 1560
 ATGGGTCTCG ATCCCGGGTT GACCGGATGC GCGCTGTCCG TCATCGAGAG TGGGCGTGGT 1620
 CGGCGCTCA CCGCGCTGGA TGTGACGTC GTGCGCACAC CGTGGGATGC GCGCTTGGCG 1680
 CAGCGCTGT TGGGATCAG CGATCGCTG GAGCACTGCC TGGACACCCA TCATCGGAG 1740
 GTGCTGGTA TCGAACGCT GTTCTCTCAG GTTAACTGA CCACGGTGAT GGGCACCGG 1800
 CAGCGCGCGG CGTCTATCCG CCGCGCGGTG GCGAAAGCTG GTGTCGACCT GCATTTCCAT 1860
 ACCCGGCGG AGGTCAAGCG CGCGGTCACT GCGAACGCTT CCGCAGACAA GGTCAAGTC 1920
 ACC 1923

1. INFORMATION FOR SEQ ID NO 191:

- A. SEQUENCE CHARACTERISTICS
- 1. LENGTH: 1055 base pairs
- 2. TYPE: nucleic acid
- 3. STRANDEDNESS: single
- 4. TOPOLOGY: linear

11. SEQUENCE DESCRIPTION: SEQ ID NO 191:

TCGATCTCG AGTCTCAAGC CGAATCGA GTTGGCTTCA AATTCTCTG CGCGCGCGGA
 100

GGGGATGGTT CAGACGTAAC GGTTGGCTAG GTCGAAACCC GCGCCAGGGC CGCTGGACGG 300
 GCTCATGGCA GCGAAATTAG AAAACCCGGG ATATTGTCCG CGGATTGTCA TACGATGCTG 360
 AGTGCTTGGT GGTTCGTGTT TAGCCATTGA GTGTGGATGT GTTGAGACCC TGGCCTGGAA 420
 GGGGACAACG TGCTTTTGCC TCTTGGTCCG CCTTTGCCGC CCGACGCGGT GGTGGCGAAA 480
 CCGGCTGAGT CCGGAATGCT CCGCGGGTTG TCGGTTCCGC TCAGCTGGGG AGTGGCTGTG 540
 CCACCCGATG ATTATGACCA CTGGGCGCCT GCGCCGGAGG ACGGCGCCGA TGTCGATGTC 600
 CAGGCGGCCG AAGGGGCGGA GCGAGAGGCC GCGGCCATGG ACGAGTGGGA TGAGTGGCAG 660
 GCGTGAACG AGTGGGTGGC GAGAACGCT GAACCTGCTT TTGAGGTGGC ACGGAGTAGC 720
 AGCAGCGTCA TTCCGCATTC TCGGCGCGCC GGCTAGGAGA GGGGGCGCAG ACTGTCGTTA 780
 TTGACCACT GATCGGCGGT CTCGGTGTTC CCGCGGCCGG CTATGACCAAC AGTCAATGTG 840
 CATGACAAGT TACAGGTATT AGGTCCAGGT TCAACAAGGA GACAGGCAAC ATGGCAACAC 900
 GTTTTATGAC GGATCCGCAC GCGATGCGGG ACATGGCGGG CCGTTTTGAG GTGCACGCCC 960
 AGACGGTGGG GGACGAGGCT GCGCGGATGT GGGCGTCCGC GCAAAACATC TCGGCGCGGG 1020
 GCTGAGTGG CATGGCCGAG GCGACCTGTC TAGAC 1055

2. INFORMATION FOR SEQ ID NO:192:

2. SEQUENCE CHARACTERISTICS

- A. LENGTH: 150 base pairs
- B. TYPE: nucleic acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

3. SEQUENCE DESCRIPTION: SEQ ID NO:192

GCGCTGCTTT GTTGGCATAA TCGGCGCGGG AGGCTGGAAG AGACTGGCC GTGCTGTGTT 1
 TCGGCTGGA GACCGCGGAT GCGGCAAC TCGGATATC GTTGGCAATG ATCGAGCTTG 120
 TCGGCTGCTT GACCGAGGCA GCGGCGGCTG TCGGACGCGG GCTGCGATGAC GTTGGTATAG 180
 AGTGGCGGTT GCGCGCGGAG GAATTTGTAG TACTGGGTGA GCGCTGCTT ACCGGGAGTG 240
 GAGAGCTTCA GGTGGAAGCT TCGGAGTGG GCGGCGAGGT AGTGAAGCG GAAACCAAAA 300
 TCGGCTGGA GCGCTGCTTT GCGGCGGAG AGGAGGAA TCGGCTGGA TCGGCTGGA

- (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:193:

```

AACGGGCCCCG TGGCCACCGC TCCTCTAAGG GCTCTCGTTG GTCGCATGAA GTGCTGGAAG      60
GATGCATCTT GGCAGATTCC CGCCAGAGCA AAACAGCCGC TAGTCCTAGT CCGAGTCGCC      120
CGCAAAGTTC CTCGAATAAC TCCGTACCCG GAGCGCCAAA CCGGGTCTCC TTCGCTAAGC      180
TGC CGGAACC ACTTGAGGTT CCGGGACTCC TTGACGTCCA GACCGATTCC TTCGAGTGGC      240
TGATCGGTTT GCGCGGCTGG CGCGAATCCG CCGCCGAGCG GGGTGATGTC AATCCAGTGG      300
CTGGCCTGGA AGAGGTGCTC TACGAGCTGT CTCGATCGA CGACTTCTCC      360
  
```

(2) INFORMATION FOR SEQ ID NO:194:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 679 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:194:

```

Glu Gln Pro Lys Gly Pro Phe Gly Glu Val Ile Gln Ala Phe Ala Asp      1
1      5      10      15
Gly Leu Ala Gly Lys Gly Lys Gln Ile Asn Thr Thr Leu Asn Ser Leu      20
20      25      30
Ser Gln Ala Leu Asn Ala Leu Asn Glu Gly Arg Gly Asp Phe Phe Ala      35
35      40      45
Val Val Arg Ser Leu Ala Leu Phe Val Asn Ala Leu His Gln Asp Asp      50
50      55      60
Gln Gln Phe Val Ala Leu Asn Lys Asn Leu Ala Gln Phe Thr Asp Arg      65
65      70      75
Leu Thr His Ser Asp Ala Asp Leu Ser Asn Ala Ile Gln Gln Phe Asp      80
80      85      90
Ser Leu Leu Ala Val Ala Arg Pro Phe Phe Ala Lys Asn Arg Gln Val      95
95      100      105      110
Leu Thr His Asp Val Asn Asn Leu Ala Ser Val Thr Thr -      115
115
  
```

185

Thr Leu Ala Ala Asn Ile Asn Gln Leu Tyr His Pro Thr His Gly Gly
 145 150 155 160
 Val Val Ser Leu Ser Ala Phe Thr Asn Phe Ala Asn Pro Met Glu Phe
 165 170 175
 Ile Cys Ser Ser Ile Gln Ala Gly Ser Arg Leu Gly Tyr Gln Glu Ser
 180 185 190
 Ala Glu Leu Cys Ala Gln Tyr Leu Ala Pro Val Leu Asp Ala Ile Lys
 195 200 205
 Phe Asn Tyr Phe Pro Phe Gly Leu Asn Val Ala Ser Thr Ala Ser Thr
 210 215 220
 Leu Pro Lys Glu Ile Ala Tyr Ser Glu Pro Arg Leu Gln Pro Pro Asn
 225 230 235 240
 Gly Tyr Lys Asp Thr Thr Val Pro Gly Ile Trp Val Pro Asp Thr Pro
 245 250 255
 Leu Ser His Arg Asn Thr Gln Pro Gly Trp Val Val Ala Pro Gly Met
 260 265 270
 Gln Gly Val Gln Val Gly Pro Ile Thr Gln Gly Leu Leu Thr Pro Glu
 275 280 285
 Ser Leu Ala Glu Leu Met Gly Gly Pro Asp Ile Ala Pro Pro Ser Ser
 290 295 300
 Gly Leu Gln Thr Pro Pro Gly Pro Pro Asn Ala Tyr Asp Glu Tyr Pro
 305 310 315 320
 Val Leu Pro Pro Ile Gly Leu Gln Ala Pro Gln Val Pro Ile Pro Pro
 325 330 335
 Pro Pro Pro Gly Pro Asp Val Ile Pro Gly Pro Val Pro Pro Val Leu
 340 345 350
 Ala Ala Ile Val Phe Pro Arg Asp Arg Pro Ala Ala Ser Glu Asn Phe
 355 360 365
 Asp Tyr Met Gly Leu Leu Leu Leu Ser Pro Gly Leu Ala Thr Phe Leu
 370 375 380
 Phe Gly Val Ser Ser Ser Pro Ala Arg Gly Thr Met Ala Asp Arg His
 385 390 395 400
 Val Leu Ile Pro Ala Ile Thr Gly Leu Ala Leu Ile Ala Ala Phe Val
 405

[illegible]

INFORMATION FOR SEQ ID NO 188

Journal of Management Education 30(6)p.789-804

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:195:

Thr Pro Glu Lys Ser Phe Val Asp Asp Leu Asp Ile Asp Ser Leu Ser
 1 5 10 15
 Met Val Glu Ile Ala Val Gln Thr Glu Asp Lys Tyr Gly Val Lys Ile
 20 25 30
 Pro Asp Glu Asp Leu Ala Gly Leu Arg Thr Val Gly Asp Val Val Ala
 35 40 45
 Tyr Ile Gln Lys Leu Glu Glu Glu Asn Pro Glu Ala Ala Gln Ala Leu
 50 55 60
 Arg Ala Lys Ile Glu Ser Glu Asn Pro Asp Ala Ala Arg Ala Asp Arg
 65 70 75 80
 Cys Val Ser Pro Thr Ser Gln Ala Arg Asp Ala Arg Arg Pro Leu Ala
 85 90 95
 Arg Ser Ala Arg Leu Ala Cys Arg Arg Leu Pro Ala Ser Val Pro Thr
 100 105 110
 Thr Arg Arg Asp Pro Arg Glu Arg
 115 120

1. INFORMATION FOR SEQ ID NO:196:

- (A) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 99 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

xi. SEQUENCE DESCRIPTION: SEQ ID NO:196:

Leu Ala Cys Gln Cys His Arg Arg Tyr Asp Val Gly Ile Gln Phe Arg
 1 5 10 15
 His Pro Ala Gly Pro Val Ala Thr Gln Ser Gly Pro Pro Gly Ser Ser
 20 25 30
 Glu Ala His Ser Arg Gln Val Arg Ala Gln Cys Gly Ala Gly Phe Leu
 35 40 45
 His Arg Arg Pro Ala Val Ser Gly Ala Leu Pro Pro Asn Asn Ala Ser
 50 55 60
 Thr Gly His Arg Ser Arg Ala Ala Asn Ser Gln Arg His Leu Leu Ala
 65 70 75 80 85 90 95

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 119 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:197:

Ala Ser Leu Leu Ala Tyr Ser Ala Ala Ala Ser Thr Ala Leu Ala
 1 5 10 15
 Val Ala Cys Val Arg Ala Asp His Arg Asp Arg Arg Thr Ile Arg Asp
 20 25 30
 His Leu Ala Met Ile His Leu Ala Gln Leu Val Thr Gln Pro Pro Gly
 35 40 45
 Gly Val Arg Gln Arg Leu His His Leu Gly Ile Ala Val Ala Pro Gln
 50 55 60
 Pro Gln Glu Val Val Val Leu Ala His His Leu Val Thr Gly Thr Gly
 65 70 75 80
 Glu Val Gln Gly Glu Gly Arg His Val Ala Ala Glu Val Val Asp Pro
 85 90 95
 Glu Asn Gln Ile Leu Arg Gln Val Leu Gly Pro Ala Pro His Asp Lys
 100 105 110
 Pro Asp Ala Gly Ile Gly Gln
 115

2 INFORMATION FOR SEQ ID NO:198:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 116 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:198:

Arg Ala Arg Gly His Arg Ser Ser Lys Gly Ser Arg Trp Ser His Glu
 1 5 10 15
 Val Leu Glu Gly Cys Ile Leu Ala Asp Ser Arg Gln Ser Lys Thr Ala
 20 25 30
 Ala Ser Pro Ser Pro Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser
 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 116

Glu Val Pro Gly Leu Leu Asp Val Gln Thr Asp Ser Phe Glu Trp Leu
 65 70 75 80
 Ile Gly Ser Pro Arg Trp Arg Glu Ser Ala Ala Glu Arg Gly Asp Val
 85 90 95
 Asn Pro Val Gly Gly Leu Glu Glu Val Leu Tyr Glu Leu Ser Pro Ile
 100 105 110
 Glu Asp Phe Ser
 115

(2) INFORMATION FOR SEQ ID NO:199:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 811 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO:199:

TGCTACGCAG CAATCGCTTT GGTGACAGAT GTGGATGCCG GCCTCGCTGC TGGCGATGGC 60
 GTGAAAGCGG CCGAGCTGTT CCGCCGATTC GGGGAGAACA TCGAACTGCT CAAAAGGCTG 120
 GTGCGGGCGG CCGATGATCG GGTGCGCGAC GAGGCGACGT GCACGCACTG TCAACACCAC 180
 GCGCGTGTTC GGTTCGCTTT CGAGCTGCCA TGAGGGTGCT GGTGACCGGC GCGGCGGGCT 240
 TCATCGGGTC GCGCGTGGAT GCGGCGTTAC GGGGTGCGGG TCACGACGTG GTGGGCGCTG 300
 ACCGCGTGTG GCGGCGCGCG CACGCGCCAA ACCGCGTGTG TCGACCGCGG TCGCAGCGGG 360
 TCGAGCTGCG CGACTCCAGC GCGCTGGCGG GGTGTGTGGC GGTGTGCGAT GTGGTGTGTC 420
 ACCAGCGCGG CATGGTGGGT GCGGCGGTCA ACCGCGCGGA CGACCGCGCG TATGGCGCGG 480
 ACAAGCATTT CCGCACCAGG GTGCTGCTGG CCGAGATGTT GCGGCGCGG GTGCGCGCTT 540
 TGGTGTGCGG GTGCTGATG GTGTTTACG CCGAGCGGCG GTATGCTGT TCGCAGCATG 600
 GAAGGCTGCA CCGCTGCGG GGTGCGCGAG CCGAGCTGCA GTATGCTGT TCGCAGCATG 660
 GTTGTGCGGG GTGCGCGGAG TCGATGATGT GCGAATGCTT GAGTGAAGAT GCGGCGGTGC 720
 GTTGTGCGAG CCGTGTACCG GCGCAGAGAG CCGCAGGAG CATTAAGCGG TCGGCTGCTG 780
 GGAAGCAAT GCGGCTTGGG TCGTGTGCTT 840

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:200:

```

GTCCCGCGAT GTGGCCGAGC ATGACTTTTCG SCAACACCCG CGTAGTAGTC GAAGATATCG      60
GACTTTGTTCG TCCCGGTGGC GGGATAGAGC ACC1GTCCGC GTTGGTCAGC GTCACCCGTT      120
GCTCGGACGC CGAACCCATG CTTTCAACGT AGCCTGTCCG TCACACAAGT CGCGAGCGTA      180
ACGTCACGGT CAAATATCGC GTGGAATTTG GCCGTGACGT TCCGCTCCCG GACAATCAAG      240
GCATACTCAC TTACATCCGA GCCATTTGGA CCGGTTTCGAT CGCCTTCGGG CTGGTGAACG      300
TCCCGGTCAA GGTGTACAGC GGTACCGGAG ACCACGACAT CAGGTTCACG CAGGTGCACG      360
CCAAGGACAA CGGACCCATC CGGTACAGG GCGTCTCCGA GCGGTCTGGG CAGGTGCTCG      420
ACTACCGCGA TCTTGGCCCG GCGTACGAGT CCGCGGACCG CCAAATGGTG GCGATCACCG      480
ACGACGACAT CGCCAGCTTG CCTGAAGAAG GCAGCCCGGA GATCGAGGTG TTGGAGTTCC      540
TCCCGCCCGC CGACGTGGAC CCGATGATGT TCGACCCGAG CTACTTTTTG GAGCCTGATT      600
CGAAGTCGTC GAAATCGTAT GTGCTGCTGG CTAAGACACT CGCCGAGACC GACCGGATGG      660
CGATCGTGGA TCGCCCGACC GCGCGTGAAT GCAGGAAAAA TAAGAGCCCG TATCCACAAT      720
TCCCGCTCGA GCTCGGCTAC CACCAACGGT AGAAGGATCG AGACATTCGG GAGCTGAAGT      780
TCCCGCTGAT AGAAGGCTGT TCCCGGATTT ATCAAACGCA AATACGCTT ACTCATGCGA      840
TCCCGCTGCG TCACCGGATG CGACGTTTTT TCCAGGCTCG ACCGCTTCGG GCGCGACCTG      900
AATGCGGCTG GCATCCGAGC GTTCCCGCGA AACCGGTTCC GCGCGGTCGG CTCATCGCTT      960
CATCTT

```

INFORMATION FOR SEQ ID NO 200:

1. SEQUENCE CHARACTERISTICS:
A. LENGTH: 960 base pairs
B. TYPE: nucleic acid
C. STRANDEDNESS: single
D. TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:200:

GTCCATCCCGT GCGAATACTT GCGAATACTT GCGAATACTT

CTCGCGCGCT GCGACCGTTG CCGCGCTTGC CGATCAACAT GCGGCTGGCG CCACCGTTGC 240
 CACCCACGCG ACCGGCTCCG CCGACCGCGC CGACACCAAG CGAGCTGCCG CCGGAGCCAC 300
 CATCACCACC TACGCCACCG ACCGCGCAGA CACCAGCGAC CGGGTCTTCG TGAAACGTCC 360
 CGGTGCCACC ACCGCGCGCG TTACCGCCAA CCCCACCGGC AACGCGGGCG CCGCCATCCC 420
 CGCGGGCCCC GCGGTGCCG CCGTTGCCG CGTTGCCGAA CAACAACCCG CCGGCGCCGC 480
 CGTTGCCGCG CCGCGCGCGG GTCCCGCGCG CCGCGCGGAC GCCAAGGCGG CTGCCGCGCT 540
 TGCGCGCATC ACCACCGTTG CCGCGGACCA CATCGGGTTC TGCCTCGGGG TCTGGGCTGT 600
 CAAACCTCGG GATGCCAGCG TTGCCCGCGC TTCCCGCGGG CCCCCCGTG GCGCGCTCAC 660
 LACCGATACC ACCCGCGCGA CCGCGCGGAC CGTTGCCGCG ATCACCGAAT AGCAACCGCG 720
 CGCGCGCACC ATTGCCCGCA GTCCCGCTG CCGCACCGTC GCGCGCGGAG GCGGCACTGG 780
 CAGCGCGCTT ACCACCGAAA CCGCGCGTAC CACCGGTAGA GSTGGCACTG GCGATGTGA 840
 CGAAAGCGCG GCGTCCGGCG CCGCGCGTAC CACCGCGACT GCGGGCGGCT ACACCGTCGG ••• 900
 ACCCGTTGCC ACCATCACC CGAAAGCGCG TCGCAATGTC GCGGTGCGCG ACTCCGCGCT 960
 CCGCGCGCTT GCGCGCGCGG CCGCGCGGAG CGCGGTAGC GCGGTACCA CCGGCACCGC 1020
 CGGTGCCCTT GCGCGAGCGT GCGGTGCGCG TGGCACCGTC GCGCGCGGTG CCACCGGTCC 1080
 CGGTGCCGCG AGTCCCATGG TCGCGCGTGC CCGGTGCGCG GCGGTGTTGA TCACCGATGC 1140
 CGGACACATC TCGCGGGCTG TCGCGGTGC TCGCGCGCGG GCGGGGCTG CGATTGACCG 1200
 CGTTTGCGCG GCGGAGGCGG GCGCGCGCGG TACCACCGCG CCGCGCATGG CCGAACAGCG 1260
 TCGGTGCGCG GCGGTACCG CCGCGACCGG CGATGCGTGC GCGCACGCTG GTGCGCGCGA 1320
 TCGGTGCGCG TCGCGCTG CCGCGACCGG TCGCGCGCTT TCGCGCGG TCGCGCGCGG 1380
 TCGGTGCGCG TCGCGCGCGG TCGGTGCGCG TCGCGCGG TCGCGCGCGG TCGCGCGCGG 1440
 TCGGTGCGCG TCGCGCGCGG TCGGTGCGCG TCGGTGCGCG TCGGTGCGCG TCGGTGCGCG 1500
 TCGGTGCGCG TCGGTGCGCG TCGGTGCGCG TCGGTGCGCG TCGGTGCGCG TCGGTGCGCG 1560
 TCGGTGCGCG TCGGTGCGCG TCGGTGCGCG TCGGTGCGCG TCGGTGCGCG TCGGTGCGCG 1620
 TCGGTGCGCG TCGGTGCGCG TCGGTGCGCG TCGGTGCGCG TCGGTGCGCG TCGGTGCGCG 1680

CCAGCAGCTG CGGCGTCGCG ATCAGCAAGG ACACCTCGCA CCTCCGGATA CCCCATATCG 1920
 CCGCACCGTG TCCCCAGCGG CCACGTGACC TTTGGTCGCT GGCTGGCGGC CCTGACTATG 1980
 GCGCGGACGG CCTTCGTTCT GATTGCCCCC GCGCGCGCAGC TTGTTGCGCG AGTTGAAGAC 2040
 GGGAGGACAG GCGGASCTTG GTGTAGACGT GGGTCAAGTG GGAATGCACG GTCCGCGGCG 2100
 AGATGAATAG GCGGACGCGG ATCTCCTTGT TGCTGAGTCC CTCACCGACC AGTAGAGCCA 2160
 CCTCAAGCTC TGTCGGTGTC AACGCGCCCC AGCCACTTGT CGGGCGTTTC CGTGACCGCG 2220
 GGCCTCGTTG CCGGTACCGG ATCGCCTCAT CGATCGATAA CGCAGTTCCT TCGGCCCAGG 2280
 CATCCTCGAA CTCGCTGTCA CCGATGGATT TTCGAAGGCT GGCTAGCGAC GAGTTACAGC 2340
 CCGCCTGGTA GATCCCCAAG CGGACCG 2367

(2) INFORMATION FOR SEQ ID NO:202:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 376 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:202:

Gln Pro Ala Gly Ala Thr Ile Ala Ala Ser Ser Pro Cys Ala Thr Val
 1 5 10 15
 Gly Ala Gly Gly Gly Thr Gly Ser Pro Val Thr Thr Glu Thr Ala Ala
 20 25 30
 Thr Thr Gly Arg Gly Gly Ser Gly Asp Val Tyr Glu Ser Ala Ala Ser
 35 40 45
 Gly Ala Ala Ala Thr Thr Ser Thr Ala Ile Gly Tyr Thr Val Gly Pro
 50 55 60
 Thr Ala Thr Ile Thr Ala Val Gly Ala Arg Asn Val Ala Leu Arg Asp
 65 70 75 80
 Ser Ala Val Ala Ala Val Ala Ala Ala Thr Gly Ser Gly Gly Thr
 85 90 95
 Ala Val Thr Thr Gly Thr Ala Gly Ser Leu Ala Ala Ala Cys Arg Arg
 100 105 110

His Ile Cys Arg Ala Val Pro Gly Ala Gly Arg Gly Ala Gly Arg Gly
 145 150 155 160
 Ile Asp Pro Val Cys Pro Gly Glu Ala Gly Ala Ala Gly Thr Thr Gly
 165 170 175
 Ala Ala Met Ala Glu Gln Pro Gly Val Ala Ala Val Thr Ala Arg Thr
 180 185 190
 Pro Asp Ala Cys Gly His Ala Gly Ala Ala Asp Thr Ala Val Ala Ala
 195 200 205
 Val Ala Pro Gln Pro Pro Pro Val Pro Thr Gly Thr Ala Gly Arg Ala
 210 215 220
 Gly Thr Thr Gly Pro Ala Val Ala Ala Val Ala Asp Gln Pro Gly Arg
 225 230 235 240
 Ala Ser Ala Ala Ala Gly Leu Thr Glu Pro Ala Ser Arg Ala Val Ala
 245 250 255
 Thr Val Ala Lys Gln Gln Pro Ala Gly Arg Ala Arg Leu Pro Gly Cys
 260 265 270
 Arg Pro Val Gly Ala Val Ser Asp Gln Arg Ala Pro Gln Lys Arg Leu
 275 280 285
 Gly Gly Arg Ile His Arg Thr Gln Gln Thr Pro Leu Asn Ser Gly Phe
 290 295 300
 Ser Ala Gly Ile Pro Thr Arg Gly Arg Ser Gln Arg Leu His Lys Leu
 305 310 315 320
 Leu Val Lys Arg Lys His Leu Tyr Ala Gln Arg Leu Ile Leu Pro Ser
 325 330 335
 Met Gly Pro Glu Gln Pro Arg Asn Arg Arg Arg His Phe Ile Gly Ser
 340 345 350
 Arg Ser His His Phe Arg Arg Ala Arg Arg Arg His Arg Ile Ser Arg
 355 360 365
 Ala His Leu Arg Thr Asn Ser Arg
 370 375

2 INFORMATION FOR SEQ ID NO:203

1 SEQUENCE CHARACTERISTICS

1.1 LENGTH: 381 base pairs

GGCCAAAACG CCGCGGCGAT CGCGGCCACC GAGGCCGCTT ACGACCAGAT GTGGGCCACG 60
GACGTGGCGG CGATGTTTGG CTACCATGCC GGGGCTTCGG CGGCCGTCTC GGCCTTGACA 120
CCGTTCGGCC AGGCGCTGCC GACCGTGGCG GCGGGCGGTG CGCTGGTCAG CGCGGCCGCG 180
GCTCAGGTGA CCACGCGGGT CTTCGCAAC CTGGGCTTGG CGAACGTCCG CGAGGGCAAC 240
GTCCGCAACG GTAATGTCCG GAATTCAAT CTCGGCTCGG CCAACATCGG CAACGGCAAC 300
ATCGGCAGCG GCAACATCGG CAGCTCCAAC ATCGGGTTTG GCAACGTGGG TCCTGGGTTG 360
ACCGCAGCGC TGAACAACAT CGGTTTCGGC AACACCGGCA GCAACAACAT CGGGTTTGGC 420
AACACCGGCA GCAACAACAT CGGGTTCGGC AATACCGGAC ACCGCAACCG AGGTATCGGG 480
CTCACGGSTA GCGGTTTGGT GGGGTTCCGG GGCCTGAACG GGGGCACCGG CAACATCGGT 540
GTGTTCAACT CGGGCACCGG AAACGTCCGG ATCGGCAACT CGGGTACCGG GAACTGGGGC 600
ATTGGCAACT CGGGCAACAG CTACAACACC GGTTTTGGCA ACTCCGCGCA CGCCAACACG 660
GGCTTTTCA ACTCCGAAT AGCCAACACC GCGGTCCGCA ACCGCGGCAA CTACAACACC 720
GGTAGCTACA ACCCGGGCAA CAGCAATACC GCGGGCTTCA ACATGGGGCA GTACAACACG 780
GGGTACCTGA ACACCGGCAA CTACAACACC GGTTCGGCAA ACTCCGGCAA TGTCAACACC 840
GGGCGCTTCA TTACTGCCAA GTTCAACAAC GGTTCCTTGT GCGCGGGCGA CCACCAAGGC 900
TTGATTTTGG GAGCGCGCGG GTTCTTCAAC TCGACCACTG CGCGCTCGTG GGGATTCTTC 960
AAGAGCGGTG CGGTATCGCG GTCCGCTTTC TTAACTCGG GTGCCAACAA TTCTGGCTTC 1020
TTCAACTCTT GTCGGGGGG CATCGSTAAG TCGGGCTCG CAAACGCGGG CGTGCTGGTA 1080
TTGGGCGTGA TCAACTCGGG CAACACCGTA TCGGGTTTGT TCAACATGAG CTTGCTGGTT 1140
ATCAACACCG GTTCTTGTAT GTGGGCTTC TTCAACCGG GAAGCAACAT GTGGGATTTT 1200
TTTGTGGGGT AAGGTTTGT CATCTCGG GTCAACAACT GCGCTCGCTT GAACATTTCTT 1260
GGTAACCGCA AATCGGCA TTACAACATT TCGGCAAGCG GAAACGTCCG TGAATTCAAC 1320
ATGCTTGGCA GCGGCAACGT GCGCACCGAA AACATCTTGG GCAGCGGCAA CGTGGCAGG 1380
TTCAACATCG GCACTGAAAA CATCGGAGTA TTCAATCTTT GTTCCGCAAG CCGTGGAAAA 1440
TAAACATCG CATCGGAAAA GTTGGGATG TAACATCTT GTTCTGAAAA CCGTGGAAAA

AACCAGCAGG GCTTCAATAT TGCTAGCGGC TGGAACTCGG GCACCGGCAG CAGCGGCCTG 1680
 TTCAATTCGG GCACCAATAA CGTTGGCATT TTCAACGCGG GCACCGGAAA CGTCGGCATT 1740
 GCAAACTCGG GCACCGGGAA CTGGGGTATC GGGAACCCGG GTACCGACAA TACCGGCATT 1800
 CTCAATGCTG GCAGCTACAA CACGGGCATT CTCAACGCGG GCGACTTCAA CACGGGCTTC 1860
 TACAACACGG GCACTACAA CACCGGGGGC TTCAACGTCG GTAACACCAA CACCGGCAAC 1920
 TTCAACGTCG GTGACACCAA TACCGGCAGC TATAACCCGG GTGACACCAA CACCGGCTTC 1980
 TTCAATCCGG GCAACGTCAA TACCGGGCGT TTGACACGG GCGACTTCAA CAATGGCTTC 2040
 TTGGTGGCGG GCGATAACCA GGGCCAGATT GCCATCGATC TCTCGGTCAC CACTCCATTG 2100
 ATCCCCATAA ACGAGCAGAT GGTGATTGAC GTACACAAAG TAATGACCTT CGGCGGCAAC 2160
 ATGATCAGCG TCACCGAGGC CTCGACGCTT TCGCCCAAAA GCTTCTATCT GAGCGGTTTT 2220
 TTCTTCTTGG GCGCGGTCAA TCTCAGCGCA TCCACGCTGA CCGTTCCGAC GATCACCCTC 2280
 ACCATCGGGG GACCGACGGT GACCGTCCC ATCAGCATTG TGGGTGCTCT GGAGAGCCCG 2340
 ACGATTACCT TCCTCAAGAT CGATCCGGCG CCGGGCATCG GAAATTCGAC CACCAACCCC 2400
 TCGTCCGGCT TCTTCAACTC GGGCACCGGT GGCACATCTG GCTTCCAAAA CGTCGGCGGG 2460
 GCGASTTCAG GCGTCTGGAA GAGTGGTTTT AGCAGCGCGA TAGGGGAATC GGGTTTCAG 2520
 AACCTCGGCT CGCTSCAGTC AGGCTGGGCG AACCTGGGCA ACTCGGTATC GGGCTTTTTT 2580
 AACACAGTA CGGTGAACCT CTCACGCGCG GCGAATGCTT CGGCGCTGAA CAACATCGGC 2640
 AACACGCTCT CCGGCTGCTT CCGGCTGCG AGCGGGACGA TTTTCAACGC GGGCCTTGCC 2700
 AACCTGGGCG AGTTGAACAT CCGGAGCGCT TCGTCCCGAA TTCGGCACGA GTTAGATACG 2760
 TTTTCAACGA TCATATCGCG TTTTGGGGC AGTCAATCAA AGGAATCGAA TCGGGGAAGC 2820
 TTAATTAAT AAACCAATC TCGGCTGCTT AT 2880

SEQUENCE INFORMATION FOR SEQ ID NO: 104

- SEQUENCE CHARACTERISTICS
- (A) LENGTH: 942 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS
 - (D) TOPOLOGY: linear

196

Met Trp Ala Gln Asp Val Ala Ala Met Phe Gly Tyr His Ala Gly Ala
 20 25 30
 Ser Ala Ala Val Ser Ala Leu Thr Pro Phe Gly Gln Ala Leu Pro Thr
 35 40 45
 Val Ala Gly Gly Gly Ala Leu Val Ser Ala Ala Ala Ala Gln Val Thr
 50 55 60
 Thr Arg Val Phe Arg Asn Leu Gly Leu Ala Asn Val Arg Glu Gly Asn
 65 70 75 80
 Val Arg Asn Gly Asn Val Arg Asn Phe Asn Leu Gly Ser Ala Asn Ile
 85 90 95
 Gly Asn Gly Asn Ile Gly Ser Gly Asn Ile Gly Ser Ser Asn Ile Gly
 100 105 110
 Phe Gly Asn Val Gly Pro Gly Leu Thr Ala Ala Leu Asn Asn Ile Gly
 115 120 125
 Phe Gly Asn Thr Gly Ser Asn Asn Ile Gly Phe Gly Asn Thr Gly Ser
 130 135 140
 Asn Asn Ile Gly Phe Gly Asn Thr Gly Asp Gly Asn Arg Gly Ile Gly
 145 150 155 160
 Leu Thr Gly Ser Gly Leu Leu Gly Phe Gly Gly Leu Asn Ser Gly Thr
 165 170 175
 Gly Asn Ile Gly Leu Phe Asn Ser Gly Thr Gly Asn Val Gly Ile Gly
 180 185 190
 Asn Ser Gly Thr Gly Asn Trp Gly Ile Gly Asn Ser Gly Asn Ser Thr
 195 200 205
 Asn Thr Gly Phe Gly Asn Ser Gly Asp Ala Asn Thr Gly Phe Phe Asn
 210 215 220
 Ser Gly Ile Ala Asn Thr Gly Val Gly Asn Ala Gly Asn Tyr Asn Thr
 225 230 235 240
 Gly Ser Tyr Asn Pro Gly Asn Ser Asn Thr Gly Gly Phe Asn Met Gly
 245 250 255
 Gln Tyr Asn Thr Gly Tyr Leu Asn Ser Gly Asn Tyr Asn Thr Gly Leu
 260 265 270
 Ala Asn Ser Gly Asn Val Asn Thr Gly Ala Phe Leu Thr Gly Asn Phe
 275 280 285

305
 Asn Ser Gly Ala Gly Ser Ala Ser Gly Phe Leu Asn Ser Gly Ala Asn
 325 330 335
 Asn Ser Gly Phe Phe Asn Ser Ser Ser Gly Ala Ile Gly Asn Ser Gly
 340 345 350
 Leu Ala Asn Ala Gly Val Leu Val Ser Gly Val Ile Asn Ser Gly Asn
 355 360 365
 Thr Val Ser Gly Leu Phe Asn Met Ser Leu Val Ala Ile Thr Thr Pro
 370 375 380
 Ala Leu Ile Ser Gly Phe Phe Asn Thr Gly Ser Asn Met Ser Gly Phe
 385 390 395 400
 Phe Gly Gly Pro Pro Val Phe Asn Leu Gly Leu Ala Asn Arg Gly Val
 405 410 415
 Val Asn Ile Leu Gly Asn Ala Asn Ile Gly Asn Tyr Asn Ile Leu Gly
 420 425 430
 Ser Gly Asn Val Gly Asp Phe Asn Ile Leu Gly Ser Gly Asn Leu Gly
 435 440 445
 Ser Gln Asn Ile Leu Gly Ser Gly Asn Val Gly Ser Phe Asn Ile Gly
 450 455 460
 Ser Gly Asn Ile Gly Val Phe Asn Val Gly Ser Gly Ser Leu Gly Asn
 465 470 475 480
 Tyr Asn Ile Gly Ser Gly Asn Leu Gly Ile Tyr Asn Ile Gly Phe Gly . . .
 485 490 495
 Asn Val Gly Asp Tyr Asn Val Gly Phe Gly Asn Ala Gly Asp Phe Asn
 500 505 510
 His Gly Phe Ala Asn Thr Gly Asn Asn Asn Ile Gly Phe Ala Asn Thr
 515 520 525
 Gly Asn Asn Asn Ile Gly Leu Gly Leu Ser Gly Asp Asn Gln Gln Gly
 530 535 540
 Phe Asn Ile Ala Ser Gly Trp Asn Ser Gly Thr Gly Asn Ser Gly Leu
 545 550 555 560
 Phe Asn Ser Gly Thr Asn Asn Val Gly Ile Phe Asn Ala Gly Thr Gly
 565 570 575

Gly Ile Leu Asn Ala Gly Asp Phe Asn Thr Gly Phe Tyr Asn Thr Gly
 610 615 620
 Ser Tyr Asn Thr Gly Gly Phe Asn Val Gly Asn Thr Asn Thr Gly Asn
 625 630 635 640
 Phe Asn Val Gly Asp Thr Asn Thr Gly Ser Tyr Asn Pro Gly Asp Thr
 645 650 655
 Asn Thr Gly Phe Phe Asn Pro Gly Asn Val Asn Thr Gly Ala Phe Asp
 660 665 670
 Thr Gly Asp Phe Asn Asn Gly Phe Leu Val Ala Gly Asp Asn Gln Gly
 675 680 685
 Gln Ile Ala Ile Asp Leu Ser Val Thr Thr Pro Phe Ile Pro Ile Asn
 690 695 700
 Glu Gln Met Val Ile Asp Val His Asn Val Met Thr Phe Gly Gly Asn
 705 710 715 720
 Met Ile Thr Val Thr Glu Ala Ser Thr Val Phe Pro Gln Thr Phe Tyr
 725 730 735
 Leu Ser Gly Leu Phe Phe Phe Gly Pro Val Asn Leu Ser Ala Ser Thr
 740 745 750
 Leu Thr Val Pro Thr Ile Thr Leu Thr Ile Gly Gly Pro Thr Val Thr
 755 760 765
 Val Pro Ile Ser Ile Val Gly Ala Leu Glu Ser Arg Thr Ile Thr Phe
 770 775 780
 Leu Lys Ile Asp Pro Ala Pro Gly Ile Gly Asn Ser Thr Thr Asn Pro
 785 790 795 800
 Ser Ser Gly Phe Phe Asn Ser Gly Thr Gly Gly Thr Ser Gly Phe Gln
 805 810 815
 Asn Val Gly Gly Gly Ser Ser Gly Val Thr Asn Ser Ile Leu Ser Ser
 820 825 830
 Ala Ile Gly Asn Ser Gly Phe Gln Asn Leu Gly Ser Leu Gln Ser Gly
 835 840 845
 Thr Ala Asn Leu Gly Asn Ser Val Ser Gly Phe Phe Asn Thr Ser Thr
 850 855 860
 Val Asn Ser Ser Thr Pro Ala Asn Val Ser Gly Leu Asn Asn Ile Ile
 865 870 875

Ala Gly Leu Ala Asn Leu Gly Gln Leu Asn Ile Gly Ser Ala Ser Cys
 900 905 910

Arg Ile Arg His Glu Leu Asp Thr Val Ser Thr Ile Ile Ser Ala Phe
 915 920 925

Cys Gly Ser Ala Ser Asp Glu Ser Asn Pro Gly Ser Val Ser Glu
 930 935 940

(2) INFORMATION FOR SEQ ID NO:205:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 53 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

XX: SEQUENCE DESCRIPTION: SEQ ID NO:205:

GGATCCATAT GGGCCATCAT CATCATCATC ACGTGATCGA CATCATCGGG ACC

53

(2) INFORMATION FOR SEQ ID NO:206:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 42 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

XX: SEQUENCE DESCRIPTION: SEQ ID NO:206:

GCTGAATTCA GGCCTGGGTT GCGCCGCGCT CATCTTGAAC GA

42

(2) INFORMATION FOR SEQ ID NO:207:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

XX: SEQUENCE DESCRIPTION: SEQ ID NO:207:

GAATGATGCA GGTTCGAAAC CAGCGAGCGG T

31

(2) INFORMATION FOR SEQ ID NO:208:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid

CTGTGAATTC AGCGCTGGAA ATCGTCCGGA T

31

(2) INFORMATION FOR SEQ ID NO:209:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:209:

GGATCCAGCG CTGAGATGAA GACCGATGCC GCT

33

(2) INFORMATION FOR SEQ ID NO:210:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 38 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:210:

GGATATCTGC AGAATTCAGG TTAAAGGCC ATTTGCCA

38

(2) INFORMATION FOR SEQ ID NO:211:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:211:

GGGATGCCGA GCGACGTGGT CACACGGCC

30

(2) INFORMATION FOR SEQ ID NO:212:

- (1) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 37 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:212:

CTTATGAAA TTCTAGGCT GTTAAGTCT TCTGCGG

37

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:213:

TGGCGAATGG GACGCGCCCT GTAGCGGCGC ATTAAGCGCG GCGGGTGTGG TGGTTACGGG	60
CAGCGTGACC GCTACACTTG CCAGCGCCCT AGCGCCCGCT CCTATCGCTT TCTTCCCTTC	120
CTTTCTCGCC ACGTTCGCGG GCTTTCCCGG TCAAGCTCTA AATCGGGGGG TCCCTTTAGG	180
GTTCTGATTT AGTGCTTTAC GGCACCTCGA CCCCCAAAAA CTTGATTAGG GTGATGGTTC	240
ACGTAGTGGG CCATCGCCCT GATAGACGGT TTTTCGCCCT TTGACGTTGG AGTCCACGTT	300
CTTTAATAGT GGACTCTTGT TCCAAACTGG AACAACTCTC AACCTATCTT CGGTCTATTC	360
TTTTGATTTA TAAGGGATTT TGCCGATTTG GGCTATTGGG TTAATAAATG AGGTGATTTA	420
ACAAAAATTT AACCCGAATT TTAACAAAAT ATTAACGTTT ACAATTTCAG GTGGCACTTT	480
TGGGGGAAAT GTGCGCGGAA CCCCTATTTC TTTATTTTTC TAAATACATT CAAATATGTA	540
TCCGCTCATG AATTAAITCT TAGAAAAACT CATCGAGCAT CAAATGAAAC TGCAATTTAT	600
TCATATCAGG ATTATCAATA CCATATTTTT GAAAAAGCGG TTTCTGTAAT GAAGGAGAAA	660
ACTCAGCGAG GCAGTTCCAT AGGATGGCAA GATCCTGGTA TCGGTCTGGG ATTCCGACTC	720
GTCCACATC AATACAACCT ATTAATTTCC CCTCTGAAA AATAAGGTTA TCAAGTGAGA	780
AATCAGCATG AGTGACGACT GAATCGGCTG AGAATGGCAA AAGTTTATGC ATTTCTTTCT	840
AGACTTCTTC AACAGGCCAG CCATTACGCT CTCATGAAA ATGCTGCGCA TCAACCAAAAT	900
CGTTATTGAT TCGTGATTCC CCTGAGCGA GACGAAATAC GCGATCGCTG TTAAGAGGAC	960
AATTACAAAC AGGAATCGAA TGCAACCGGC TGAGGAACAG TCCAGTACA TCAACAATAT	1020
TTTAATGCA ATCAGGATAT TCTTCTAATA CTGTAATGC TCTTTTCTCT GCGATCGCAG	1080
CTTTAAATTA CATCGCATCA CAGCAATAT TTAATAATG TTTATGCTT TGAAGAGGCA	1140
TAAATTTCT CAGCCAGTTT AGTCTGACA TCTTATCTCT AACATGATTC GCAACGCTAC	1200
TTTTTCTAT TTTGAGAAAC AACTCTGGCG CATCGGCTTT CCGATACAAAT CGATAGATTC	1260
TGGCACTTAA TTGCCCCGAA TTATCGCGAG CCGATTATA CCGATATAAA TTAGCATGCA	1320
TCTTAAATCT CAATCGCCCT TTAGAGCAAT AGTTTCTTCT TTAATTTT	

GATCCTTTT TTCTGCGCGT AATCTGCTGC TTGCAAACAA AAAAACCACC GCTACCAGCG 1560
GTGGTTTGTG TGCCGGATCA AGAGCTACCA ACTCTTTTTC CGAAGGTAAC TGGCTTCAGC 1620
AGAGCGCAGA TACCAAATAC TGTCTTCTA GTGTAGCCGT AGTTAGGCCA CCACTTCAAG 1680
AACTCTGTAG CACCGCCTAC ATACCTCGCT CTGCTAATCC TGTTACCAGT GGCTGCTGCC 1740
AGTGGCGATA AGTCTGTCT TACCGGGTTG GACTCAAGAC GATAGTTACC GGATAAGGCG 1800
CAGCGGTCGG GCTGAACGGG GGGTTCTGTC ACACAGCCCA GCTTGAGCG AACGACCTAC 1860
ACCGAACTGA GATACCTACA GCCTGAGCTA TGAGAAAGCG CCACGCTTCC CGAAGGGAGA 1920
AAGGCGGACA GGTATCCGGT AAGCGGCAGG GTCCGAACAG GAGAGCGCAC GAGGGAGCTT 1980
CCAGGGGAAA ACCCCTGGTA TCTTTATAGT CCGTCCGGT TTCCGCACT CTGACTTGAG 2040
CGTCAATTT TGTGATGCTC GTGAGGGGGG CGGAGCCTAT GGAAAAACGC CAGCAACGCG 2100
GCCTTTTAC GGTTCCTGGC CTTTCTGCTG CTTTCTGCTC ACATGTTCTT TCCTGCGTTA 2160
TCCCTGATT CTGTGGATAA CCTATTACC GCCTTGACT GAGCTGATAC CGCTCGCCGC 2220
ACCGAACA GCGAGCGCAG CGAGTCACTG AGCGAGGAAG CGGAAGAGCG CCGATGCGG 2280
TATTTCTCC TTACCATCT GTCCGCTATT TCACACCGCA TATATGGTGC ACTCTCAGTA 2340
CAATCTGCTC TGATGCGCA TAGTTAAGCG AGTATACACT CCGCTATCGC TACGTGACTG 2400
GCTGCTGCTC TCGGCGCGAC ACCGCTCAAC ACCGCTGAC GCGCCTGAC GGGCTTCTCT 2460
CTGCGCGCA TCGCTTACA CAGAGCTCT CAGCTGCTC CGGAGCTCA TGTGTGAG 2520
CTTTCTGCGC TGTATGCGCA AACCGCGCAG CAGCTGCGG TAAAGCTCT CAGCTGCTC 2580
CTGAAGCGAT TCACAGATCT CTGCTGCTC ATCGCGCTC AGCTGCTTCA GTTCTGCGC 2640
AACCTTAT GTCTGCTT TGTAAAGCG CAGCTTCA AGGCGCTT TCTGCTT 2700
CTGCTGCTC TGTGCTT TGTGCTT TGTGCTT TGTGCTT TGTGCTT 2760
CTGCTGCTC TGTGCTT TGTGCTT TGTGCTT TGTGCTT TGTGCTT 2820
CTGCTGCTC TGTGCTT TGTGCTT TGTGCTT TGTGCTT TGTGCTT 2880
CTGCTGCTC TGTGCTT TGTGCTT TGTGCTT TGTGCTT TGTGCTT 2940
CTGCTGCTC TGTGCTT TGTGCTT TGTGCTT TGTGCTT TGTGCTT 3000

CATGCCGGCG ATAATGGCCT GCTTCTCCCG GAAACGTTTG GTGGCGGGAC CAGTGACGAA 3240
GGCTTGAGCG AGGGCGTGCA AGATTCCGAA TACCGCAAGC GACAGGCCGA TCATCGTCGC 3300
GCTCCAGCGA AAGCGGTCCT CGCCGAAAAT GACCCAGAGC GCTGCCGGCA COTGTCTAC 3360
GAGTTGCATG ATAAAGAAGA CAGTCATAAG TCGGGCGACG ATAGTCATGC CCCGCGCCCA 3420
CUGGAAGGAG CTGACTGGGT TGAAGGCTCT CAAGGGCCTC GGTCCGAGATC CCGGTGCCTA 3480
ATGAGTGAGC TAACCTACAT TAATTGCCCT GCGCTCACTG CCGCTTTCC AGTCGGGAAA 3540
CCTGTCTGCG CAGCTGCATT AATGAATCGG CCAACGCCCG GGGAGAGGCG GTTTGCGTAT 3600
TGGGCGCCAG GGTGGTTTTT GTTTTCACCA GTGAGACGGG CAACGCTCA TTGCCCTTCA 3660
CCSCCTGGCG CTGAGAGACT TCCAGCAAGC GGTCCACGCT GGTTCGCCCC AGCAGGCCAA 3720
AATCCTGTTT GATGGTGGTT AACGGCGGGA TATAACATGA GCTGTCTTCC GTATCGTCCT 3780
ATCCCACTAC CGAGATATCC GCACCAACCG GCACCCCGGA CTGGTAATG GCGCGCATTC 3840
GCCCCAGCGC CATCTGATCG TTGGCAACCA GCATCCGAGT GGGAAAGATG CCGTCATTCA 3900
GCATTTTGCAT GGTTCCTTGA AAACCGGACA TGGCACTCCA GTCCCTTTCC CGTTCCGCTA 3960
TCCCTGGAAT TTGATTGCGA GTGAGATATT TATCCAGCCG AGCCAGACCG AGACCGCCCG 4020
AGACAGAACT TAATGGGCGC GCTAACAGCG CGATTGCTG GTGACCCCAAT GCGACAGAT 4080
GCTCCACCGC CAGTCCCTTA CCGTCTTCTT GGGAGAAAAT AATACTGTTG ATGGGTGTCT 4140
GTTCAGAGAC ATCAGAAAAT AAACCGGAA ATTAGTCCA CCGAGCTTCC ACAGCAATGG 4200
CATCTGTGTC ATCCAGCGGA CAGTTAATGA TCAGCCCACT GACGCTTTC GCGAGAAGAT 4260
TGTGCACCGC CCGTTTACAG CTTTCGACCG CCGTCTTTC TACCATCGAC ACCACCGCC 4320
TCCAGCGCA TTGATCGCGT CAGATTTA TTCCCGGAC AATTTCGAC GCGCTTCCA 4380
GGCTGCAAT CAGGCTGCA ATCAGATCA CAGCAATTC TTTCGCGCG AGTTCTTCT 4440
CCAGCTTCTT GCGAATGTAA CTGAGTCCG CAGTCCCGG TTCCACTTTT TCGCGCTTT 4500
TCCAGAAAAC GTGGCTGGCG TCGTTCGCA CCGCGAAAAC GGTCTGATAA GAGACACCG 4560
CATACTGTA GACATCTAT AACCTACTG GTTCAGACT CACCACTCTG AATTGACTCT 4620
CTTCGCGCGT CATCTGCG ATACAGGAA AGCTTTTCTT CATTCTATG GTTCGCGTA 4680

CCGGCCACGG GGGCTGCCAC CATACCCACG CCGAAACAAG CGCTCATGAG CCGGAAGTGG 4860
CGAGCCCGAT CTTCGCCATC GGTGATGTGG GCGATATAGG CGCCAGCAAC CGCACCTGTG 4920
GCGCCCGTGA TGCCGGCCAC GATCCGTCCG GCGTAGAGCA TCGAGATCTC GATCCCGCGA 4980
AATTAATACG ACTCACTATA GGGGAATTGT GAGCGGATAA CAATTCGCCCT CTAGAAATAA 5040
TTTTGTTTAA GTTTAAGAAG GAGATATACA TATGGGCCAT CATCATCATC ATCACGTGAT 5100
CGACATCATC GGGACCAGCC CCACATCCTG GGAACAGGCG CCGGCGGAGG CGGTCCAGCG 5160
GGCCCGGGAT AGCGTCGATG ACATCCCGCT CGCTCGGGTC ATTGAGCAGG ACATGGCCGT 5220
GGACAGCCCC GCGAAGATCA CCTACCGCAT CAAGCTCGAA GTGTCTTCA AGATGAGGCC 5280
JUCGCAACCG AGGGGCTGCA AACCACCGAG CGGTTCGCTT GAAACGGGCG CCGGCGCCCG 5340
TACTGTCCCG ACTACCCCGG CGTCGTCCCG GGTGACGTTG GCGGAGACCG GTAGCAGCCT 5400
GCTGTACCCG CTGTTCAACC TGTGGGGTCC GGCCTTTTAC GAGAGGTATC CGAACGTGAC 5460
GATCACCCCT CAGGSCACCG GTTCTGGTGC CGGGATCGCG CAGGCCCGCG CCGGGACCGT 5520
CAACATTGGG GCGTCGACG CCTATCTGTT GGAAGGTGAT ATGGCCCGCG ACAAGGGGCT 5580
GATGAACATC CGCTAGCCA TCTCGGTCA GAGGTCAAC TACAACCTGC CCGGAGTGAG 5640
CGAGCAGCTC AAGCTGAACG GAAAGTCTCT GCGGGCCATG TACCAGGGCA CCATCAAAAC 5700
GTGGGACGAC CCGCAGATCG GTGCGCTCAA GCGGCGCTG AACCTGCGCG GCACCCCGGT 5760
AGTTCCGCTG CACCGTCCG ACGGCTCCG TGACGCTTC TTGTTGAGG AGTACCTGTC 5820
CAAGCAGAT CCGAGGGCT GGGGCGCTC GCGGCTTC GGCACCAAC TCGACTTCCC 5880
GGCGGTGCCG GGTGCGCTCG GTGAGAACCG CAACGGCGCG ATGCTGACCG GTTCCGCGCA 5940
GACACCGGGT TCGTCCGCT ATATCGCAT GAGCTTCTC TACGAGTCA GTGACCGGG 6000
GTTGCGCAG TCGCACTAG GATATGCTC TCGCACTCT GTTTCGCGG AGCGGCAAG 6060
ATTGAGGGC GCGGCGCTG GTTTCGATC GAAAGTCTC GCGACCAAG GATTTTCGAT 6120
GATGACCGG CCGGCGCGCG ACGGCTACCG GATCATCAAG TACGAGTAC GATCTCTCAA 6180
GAGCGCGAA AAGGAGCGCG GCACCGCGCA GAGCTTCAG GATTTCTCG ATTGCGCGAT 6240
GAGGACCGG AAGGAGCTC GTTTCGCTA GAGCTTCTC TCGAGCTCT TGACCGCGCG 6300

GGGGACGGCC GCGCAGGCGG CGGTGGTGGC CTTCCAAGAA GCAGCCAAATA AGCAGAAGCA 6540
 GGAACCTGAC GAGATCTCGA CGAATATTCG TCAGGCCCGG GTCCAATACT CGAGGGCCGA 6600
 CGAGGAGCAG CAGCAGGCGC TGTCCTCGCA AATGGGCTTT GTGCCACAA CGGCCGCTC 6660
 GCGCGCGTCG ACCCCTGCAG CGCCACCCGC ACCGGCGACA CTTTGTGCCC CCCCACCACC 6720
 GGCGCGCGCC AACACGCCGA ATGCCAGCC GGGCGATCCC AACGCAGCAC CTCCGCCGGC 6780
 GCACCCGAAC GCACCGCGCG CACCTGTGAT TGCCCCAAAC GCACCCCAAC CTGTCCGGAT 6840
 CGACAACCCG GTTGGAGGAT TCAGCTTCGC GCTGCTGCT GGCTGGGTGG AGTCTGACGC 6900
 CCGCCACTTC GACTACGGTT CAGCACTGCT CAGCAAAACC ACCGGGGACC CGCCATTTCG 6960
 CGGACAGCCG CCGCGCGTGG CCAATGACAC CGGTATCGTG CTCGCGCGGC TAGACCAAAA 7020
 GCTTTACGCC AGCGCCGAAG CCACCGACTC CAAGGCCCGG GCGCGGTTGG GCTCGGACAT 7080
 CGGTGACTTC TATATGCGCT ACCCGGGCAC CCGGATCAAC CAGGAAACCG TCTCGCTTGA 7140
 TGCCAACCGG GTGTCTGGAA GCGCGTCTTA TTACGAAGTC AAGTTCAGCG ATCCGAGTAA 7200
 TCGGAACCGG CAGATCTGGA CCGCGCTAAT CGGCTCGCCC GCGCGGAACG CACCGGACCG 7260
 CCGGCGCGCT CAGCGCTGGT TTCTGGTATG GCTCGGGACC GCGCAACAAC CGGTGGACAA 7320
 GGGCGCGGCC AAGCGCGTGG CCAATCGAT CCGGCTTTG GTGCGCGCGC CGCGCGCGCC 7380
 GGCACCGGCT CCGCAGAGC CCGCTCGCGC GCGCGCGCGC CCGCGGGAAG TCGCTCTAC 7440
 CCGGACGACA CCGACACCGT AGCGGACCTT ACCGCGCTGA GAATTTGGA TATATCCAT 7500
 ACACTGGCGG CCGCTCGAGC AGCACCACCA CCGCAGTGA GATCGCGCTG CTACAAAGC 7560
 CCGAAAGGAA GTCAGTTGG CTCTCGCAT TCTTACGAA TAAATATAT AACCGCTTG 7620
 GCTTTTAAA GCTTTTGA GCGTTTTT TCTGAAGGA TAAATATAT CCGGAT 7680

INFORMATION FOR SEQ ID NO.11:

SEQUENCE CHARACTERISTICS

- A LENGTH: 802 amino acids
- P TYPE: amino acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

11 SEQUENCE DESCRIPTION: SEQ ID NO.11

206

20 25 30
 Asp Ser Val Asp Asp Ile Arg Val Ala Arg Val Ile Glu Gln Asp Met
 35 40 45
 Ala Val Asp Ser Ala Gly Lys Ile Thr Tyr Arg Ile Lys Leu Glu Val
 50 55 60
 Ser Phe Lys Met Arg Pro Ala Gln Pro Arg Gly Ser Lys Pro Pro Ser
 65 70 75 80
 Gly Ser Pro Glu Thr Gly Ala Gly Ala Gly Thr Val Ala Thr Thr Pro
 85 90 95
 Ala Ser Ser Pro Val Thr Leu Ala Glu Thr Gly Ser Thr Leu Leu Tyr
 100 105 110
 Pro Leu Phe Asn Leu Trp Gly Pro Ala Phe His Glu Arg Tyr Pro Asn
 115 120 125
 Val Thr Ile Thr Ala Gln Gly Thr Gly Ser Gly Ala Gly Ile Ala Gln
 130 135 140
 Ala Ala Ala Gly Thr Val Asn Ile Gly Ala Ser Asp Ala Tyr Leu Ser
 145 150 155 160
 Glu Gly Asp Met Ala Ala His Lys Gly Leu Met Asn Ile Ala Leu Ala
 165 170 175
 Ile Ser Ala Gln Gln Val Asn Tyr Asn Leu Pro Gly Val Ser Glu His
 180 185 190
 Leu Lys Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile
 195 200 205
 Lys Thr Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn
 210 215 220
 Leu Pro Gly Thr Ala Val Val Pro Leu His Arg Ser Asn Gly Ser Gly
 225 230 235 240
 Asp Thr Phe Leu Thr Thr Gln Pro Ser Lys Gln Asp Pro Glu Gly
 245 250 255
 Trp Gly Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val
 260 265 270
 Pro Gly Ala Leu Gly Gln Asn Gly Asn Gly Gly Met Val Thr Gly Lys
 275 280 285 290
 Ala Gly Thr

Gly Asn Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala
325 330 335

Gly Phe Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp
340 345 350

Gly Pro Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile
355 360 365

Val Asn Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala
370 375 380

Phe Leu His Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp
385 390 395 400

Gln Val His Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp
405 410 415

Ala Leu Ile Ala Thr Ile Ser Ser Ala Glu Met Lys Thr Asp Ala Ala
420 425 430

Thr Leu Ala Gln Glu Ala Gly Asn Phe Glu Arg Ile Ser Gly Asp Leu
435 440 445

Lys Thr Gln Ile Asp Gln Val Glu Ser Thr Ala Gly Ser Leu Gln Gly
450 455 460

Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln Ala Ala Val Val Arg
465 470 475 480

Phe Gln Glu Ala Ala Asn Lys Gln Lys Gln Glu Leu Asp Glu Ile Ser
485 490 495

Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg Ala Asp Glu Glu
500 505 510

Gln Gln Gln Ala Leu Ser Ser Gln Met Gly Phe Val Pro Thr Thr Ala
515 520 525

Ala Ser Pro Pro Ser Thr Ala Ala Ala Pro Ala Pro Ala Thr Pro
530 535

Val Ala Pro Pro Pro Pro Ala Ala Ala Asn Thr Pro Asn Ala Gln Pro
545 550 555 560

Gly Asp Pro Asn Ala Ala Pro Pro Pro Ala Asp Pro Asn Ala Pro Pro
565 570 575

Pro Pro Val Ile Ala Pro Asn Ala Pro Gln Pro Val Arg Thr Thr
580 585

Asp Ala Ala His Phe Asp Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr
 610 615 620
 Gly Asp Pro Pro Phe Pro Gly Gln Pro Pro Pro Val Ala Asn Asp Thr
 625 630 635 640
 Arg Ile Val Leu Gly Arg Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu
 645 650 655
 Ala Thr Asp Ser Lys Ala Ala Ala Arg Leu Gly Ser Asp Met Gly Glu
 660 665 670
 Phe Tyr Met Pro Tyr Pro Gly Thr Arg Ile Asn Gln Glu Thr Val Ser
 675 680 685
 Leu Asp Ala Asn Gly Val Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys
 690 695 700
 Phe Ser Asp Pro Ser Lys Pro Asn Gly Gln Ile Trp Thr Gly Val Ile
 705 710 715 720
 Gly Ser Pro Ala Ala Asn Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp
 725 730 735
 Phe Val Val Trp Leu Gly Thr Ala Asn Asn Pro Val Asp Lys Gly Ala
 740 745 750
 Ala Lys Ala Leu Ala Glu Ser Ile Arg Pro Leu Val Ala Pro Pro Pro
 755 760 765
 Ala Pro Ala Pro Ala Pro Ala Glu Pro Ala Pro Ala Pro Ala
 770 775 780
 Gly Glu Val Ala Pro Thr Pro Thr Thr Pro Thr Pro Gln Arg Thr Leu
 785 790 795 800
 Pro Ala

INFORMATION FOR SEC ID NO. 115.

SEQUENCE CHARACTERISTICS

- LENGTH: 454 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

11 MOLECULE TYPE: Genomic DNA

41 SEQUENCE DESCRIPTION: SEQ ID NO: 1

2709 18GUCG TGGCGCTAC

CCGATCGCGA TCAGCTGCTT ACCGACCGGC GGGTGAACCA CCAGGCCGTA CCCGGGGTTG 360
 TCTTCCACCC CATGGTTGTT CAGCACCTGC CAGGCCTGGC GGTGCGTAAT GCTTCTCGTC 420
 GAAGATGGGG GTGCCGGCAT CCGTCACCGA GCCC 454

(2) INFORMATION FOR SEQ ID NO:216:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 470 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:216:

TCCAGAAGTA CGGCGGATCC TCGGTGGCCG AC3CCGAACG GATTCCGCCG GTCCGCCAAG 60
 GCATCGTCCG CACCAAGAAG CAAGGCAATG ACCTCGTCGT CTTCTGTCTT GCCATGGGGG 120
 ATACCACCGA CGACCTGCTG GATCTGGCTG AGCAGGTGTG CCGCGCCCGG CCGCCTCGGG 180
 AGCTGGACAT GCTGCTTACC GCCGGTGAAC GCATCTCGAA TCGGTTGGTG GCCATGGCCA 240
 TCGAGTCGCT CGGCGCGCAT GCCCGGTCTT TCACCGGTTT GCAGGCCGGG GTGATCACCA 300
 CCGGCACCCA CGGCAACGCC AAGATCATCG ACGTCACGCC GGGGCGGGTG CAAACCGCCC 360
 TTGAGGAAGG GCGGGTCGTC TTGGTGGCCG GATTCCAAGG GGTGAGCCAG GACACCAAGG 420
 ATGTCACGAC GTTGGGCCCG GCGCGCTCCG ACACCCACCG CGTCGCCATG 470

(2) INFORMATION FOR SEQ ID NO:217:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 279 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:217:

TGCGGGCGTA CCGGGCCCGG ACAAACAACG ATCGATTGAT ATCGATGAGA SAGCGAGGAA 60
 TCTTGGGCTT TCTCCAGTTG ACCGACGAGT AGCGCGCGGG CGCCTTGGAG AAGGCTGCTG 120
 TCTACGCTCG AGCGCGAGCA GAGCTCAAGG ATCGGCTGAA TCTTGGCGGC ACCAAGCTCA 180
 TCTAGTCTCT CAAGGACCGG GAGAGCGAT GAGTCTTGGG TAAATGAAG TTGTCTGCTT 240
 TCTTCTAGGC TCTGCGAAG GTGGGAAAG TCCAGGCGG 300

(2) INFORMATION FOR SEQ ID NO:218:

(i) SEQUENCE CHARACTERISTICS:

- A. LENGTH: 219 base pairs
 B. TYPE: nucleic acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

ACACGGTCCA	ACTCGACGAG	CCCCCTGCTGG	AGGTGTGCGAC	CGACAAGGTC	GACACCGAAA	60
TCCCTCGCCG	GCCSCGGGTG	TGCTGACCAA	GATCATCGCC	CAAGAAGATG	ACACGGTCCA	120
GGTCGGCGGC	GAGCTCTCTG	TCATTGGCGA	CGCCCATGAT	GCCGGCGAGG	CCGCGGTCCC	180
GGCACCCGAG	AAAGTCTCTG	CCGGCCCCAAC	CCGAATCCA			219

(2) INFORMATION FOR SEQ ID NO:219:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 342 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:219:

TGCTGCGCGA	GATCGGCGCC	GCCGCGCGCC	CCAAGCCCGC	ACCCAAGCCC	GTCCCCGAGC	60
CAGCGCCGAC	GCCGAAGGCC	GAACCCCGAC	CATCGCCCGC	GGCGGCCGAG	CCAGCCCGTG	120
CGGCCGAGGG	CGCACCGTAC	GTGAACCCGC	TGGTGCGAAA	GCTGGCGTCG	GAAAACAACA	180
TGGACCTCGC	CGGGGTGACC	GGCACCGGAG	TGGGTGGTCC	CATCCGCAAA	CAGGATGTGC	240
TGGCCGCGGC	TGAACAAAAG	AAGCGGCGGA	AAGCACCGGC	GCCGGCCGCC	CAGGCGCGCC	300
CCGCGCCGCG	CCCGAAAGCG	CCGCTGAAG	ATCCGATGCC	GC		342

(2) INFORMATION FOR SEQ ID NO:220:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 515 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:220:

GGTCTTGGT	CAGTATCAGC	GCCGACGAGG	ACGCCACCGT	GCCCGTCGGC	GGCGAGTTGG	60
TTGGGATCGG	TGTGGCTGCC	GACATGCGTG	CGCGCGCGCG	CCCCAAGCCC	GCACCCCAAGC	120
TTGTCTCTCGA	GCCAGCGCCG	ACCCCGAAGG	CGGAACCGGC	ACGATGCGCG	CGCGGGCGCG	180
ATGCACTGGG	TGGGCGCGAG	GCTGCACCGT	ACGTGACCGC	GCTGGTGCGA	AAGCTGGCGT	240
AGGAAAACAA	ATCGACCTTC	TCGGCGCTGA	CGCGCACCGG	AGTGGGTGGT	CGCATCGCGA	300
AAGAGGATGT	GCTGGCGCGG	GCTGAACAAA	AGAAGCGCGC	GAAAGCACCG	CGCGCGTGA	360
TCCTTCATCA	CCCGTTAAG	CACTTTCGCT	CAGAAGCGCG	CTTGCACCTC	TTGGCGGGTC	420
TTGGTGGCGT	CGAGGCGGTC	CGGAGCGGAG	TTAGGCTTAC	GCGGCGGAAA	TCTTCGAGTT	480
CGCGAGGAAG	GCGAUCGCGA	AGAGGCTCGC	CACCC			515

(2) INFORMATION FOR SEQ ID NO:221:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 515 base pairs
- (B) TYPE: nucleic acid

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:221:

CCGACCCCAA	GGTGCAGATT	CAACAGGCCA	TTGAGGAAGC	ACAGCGCACC	CACCAAGCGC	60
TGACTCAACA	GGCGGCGCAA	GTGATCGGTA	ACCAGCGTCA	ATTGGAGATG	CGACTCAACC	120
GACAGCTGGC	GGACATCGAA	AAGCTTCAGG	TCAATGTGCG	CCAAGCCCTG	ACGCTGGCCG	180
ACCAGGCCAC	CGCCGCCGGA	GACGCTGCCA	AGGCCACCGA	ATACAACAAC	GCCGCCGAGG	240
CGTTCGCAGC	CCAGCTGGTG	ACCGCCGAGC	AGAGCGTCGA	AGACCTCAAG	ACGCTGCATG	300
ACCAGGCGCT	TAGCGCCGCA	GCTCAGGCCA	AGAAGGCCGT	CGAACGAAAT	GCGATGGTGC	360
TGCAGCAGAA	GATCGCCGAG	CGAACCAAGC	TGCTCAGCCA	GCTCGAGCAG	GCGAAGATGC	420
AGGAGCAGGT	CAGCGCATCG	TTGCGGTCGA	TGAGTGAGCT	CGCCGCGCCA	GGCAACACGC	480
CGAGCCTCGA	CGAGGTGCGC	GACAAGATCG	AGCGTCGCTA	CGCCAACGCG	ATCGGTTCCG	540
CTGAACCTTG	CGAGAGT					557

(2) INFORMATION FOR SEQ ID NO:222:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 223 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:222:

CAGGATAGGT	TTCCACATCC	ACCTGGGTTG	CGCAGCCCGT	GCSCGACCCT	GTGATAGGCC	60
AGAGGTGGAC	CTGCCCCGAC	CGACGATCGA	TGASGAGTC	AACAGAAATG	GCCTTCTCCG	120
TCCAGATGCC	GGCACTCGGT	GAGAGCGTCA	CGAGGGGAC	CGTTACCCGC	TGGCTCAAAC	180
AGGAAGGCCA	CACGCTCGAA	CTCGACGAGT	CGCTCTGGA	GGT		223

(2) INFORMATION FOR SEQ ID NO:223:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 579 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:223:

AAGAAGTACA	TCTGCGGTT	GAGTCTGGCG	AAGCAAGGGA	CTCAAGCGGC	GCAGTAGCCG	60
ACCAGGACCA	CGGCATAAGG	CGAGTCCCGG	CGCACAAGCA	TAGGCGAGCC	CGGCTATGCC	120
AGGACTGGCA	CGGCGAGTGA	CGACATCGCG	JCGGTGCGCA	TCAGCATCTC	GGCCTTGACG	180
CAGCACTGTG	CGGCGCAGCT	TGCAACCTCT	TCTGCGTCCA	TGCGCTAGAG	CACCGGCCCG	240
AACGACATGG	JCCAGGTCCA	JGCTTTGGAT	TGCAAGGGGT	GGTAGTTGCG	TGCGGAATTC	300
GTGAGGCCCC	CGTGGAAAGT	GAAGCTTTTG	GGCTGTATTC	CTTAGAGGGA	TGCGACCGCG	360
TGCGGACCGG	GAACCAAGGA	CTTGAAGATG	ACCGCTTGA	CGAAGCATG	CGGATCGATC	420
JCGCTCTCCG	ACCGCAAGCA	CTTGAAGATG	CTTGAAGATG	CTTGAAGATG	CTTGAAGATG	480

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 484 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:224:

ACAACGATCG ATTGATATCG ATGAGAGACG GAGGAATCGT GGCCCTTCCC CAGTTGACCG	60
ACGAGCAGCG CCGGGCCCGG TTGGAGAAGG CTGCTGCCGC ACCTCGAGCG CGAGCAGAGC	120
TCAAGGATCG GCTCAAGCGT GGCGGCACCA ACCTCACCCA GGTCTCAAG GACGCGGAGA	180
GCGATGAAGT CTTGGGCAAA ATGAAGGTGT CTGCGCTGCT TGAGGCCTTG CCAAAGGTGG	240
GCAAGGTCAA GGCGCAGGAG ATCATGACCG AGCTGGAAAT TGCGGCCAC CCGCGCGCT	300
TCGTGGCCTC GGTCAACGTC AGGCGAAGGC CCTGCTGGAA AAGTTCGGCT CCGCCTAACC	360
CCCGCGGCGG ACGATGCGGG CCGGAAGGCC TGTGTTGGGC GTACCCCGCG ATACGGGGGA	420
GAAGCGGCCT GACAGGGCCA GGTCAAAATT CAGGCCGAAC GCGCCGCTGG GGGGGAACCC	480
GCCC	484

(i) INFORMATION FOR SEQ ID NO:225:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 537 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:225:

AGGACTGGCA CCGCCAGCCA CCACATCCCG TGGGTGCCGA CCAGCATCTC GGCTTGACG	60
CAGACTGTG CCGCCGAGCG TCGAACGTCT TCGTGGTCCA TGGCGTACAG CACCGGCCCG	120
AACGACATGG CCGAGGTCCA CCGTTTGGAT TCCCAAGGGT GGTAGTTGCC TCGGGAATTC	180
GTGAGGCGCG CGTGGAAAGT GAACGCTTTG CCGGTGTAGT CCGAGAGCGA GCGCACGCGG	240
TGGGGCAGCG GAACAACCGA GTTGGGACCG ACCGCTTGAC CGACCGCATG CCGATCGATC	300
CCGCTCTCGG ACGCGAACCA CGGAGCGTAG CTGGCCAGAT AGACCGCGAA TGGGATCAAC	360
CCGAGCGCAT ACCCGCTGGG AAGGACCTCA CGCGGCACTG TCCCGAGCCA CCGTCTTTGC	420
ACTTGTACT JACGTGCGCG TCCAGCTCG AAGGTCAGCG CATTGCGCGG GAAGAACAGC	480
ACGAATACA CCGCGGACCA TTGCTGCTG GAAGCTAATG CCAAGCGCA CCGCGGCG	537

(i) INFORMATION FOR SEQ ID NO:226:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 135 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

213

Gly Gly Ala Ala Ala Gly Gln Gln Ser Asp Val His Pro Phe Ala Asn
 1 5 10 15
 Leu Ile Ala Val Asp Asp Glu Arg Ala Glu Arg Arg Asp Asp Glu Glu
 20 25 30
 Arg Gln Glu Ala Val Gln Gln Arg Gly Pro Arg Gly Asp Glu Ala Asp
 35 40 45
 Pro Val Ala Asp Gln Gln His Pro Gly Asp Gly Ala Asp Gln Cys Arg
 50 55 60
 Pro Ala Asp Pro Pro His Asp Pro His His Gln Arg His Gln Asp His
 65 70 75 80
 Thr Gln Gln Gly Ala Gly Glu Pro Pro Ala Glu Ser Val Val Thr Glu
 85 90 95
 Asp Gly Leu Pro Asp Arg Asp Gln Leu Leu Thr Asp Arg Arg Val Asn
 100 105 110
 His Gln Ala Val Pro Gly Val Val Phe His Pro Met Val Val Gln His
 115 120 125
 Leu Pro Gly Leu Ala Val Arg
 130 135

(2) INFORMATION FOR SEQ ID NO:227:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 156 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(12) SEQUENCE DESCRIPTION: SEQ ID NO:227:

Gln Lys Tyr Gly Gly Ser Ser Val Ala Asp Ala Glu Arg Ile Arg Arg
 1 5 10 15
 Val Ala Glu Arg Ile Val Ala Thr Lys Lys Gln Gly Asn Asp Val Val
 20 25 30
 Val Val Val Ser Ala Met Gly Asp Thr Thr Asp Asp Leu Leu Asp Leu
 35 40 45
 Ala Gln Gln Val Cys Pro Ala Pro Pro Pro Arg Glu Leu Asp Met Leu
 50 55 60
 Leu Thr Ala Gly Gln Arg Ile Ser Asn Ala Leu Val Ala Met Ala Ile
 65 70 75 80
 Gln Ser Leu Gly Ala His Ala Arg Ser Phe Thr Gly Ser Gln Ala Gly
 85 90 95
 Val Ile Thr Thr Gly Thr His Gly Asn Ala Lys Ile Ile Asp Val Thr
 100 105 110
 Pro Gly Arg Leu Gln Thr Ala Leu Glu Gln Gly Arg Val Val Leu Val
 115 120 125
 Ala Gly Phe Gln Gly Val Ser Gln Asp Thr Lys Asp Val Thr Thr Leu
 130 135 140
 Gly Arg Gly Gln Ser Asp Thr Thr Ala Val Ala Met
 145

214

- (A) LENGTH: 92 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:228:

```

Pro Ala Tyr Pro Ala Gly Thr Asn Asn Asp Arg Leu Ile Ser Met Arg
 1           5           10           15
Asp Gly Gly Ile Val Ala Leu Pro Gln Leu Thr Asp Glu Gln Arg Ala
      20           25           30
Ala Ala Leu Glu Lys Ala Ala Ala Arg Arg Ala Arg Ala Glu Leu
      35           40           45
Lys Asp Arg Leu Lys Arg Gly Gly Thr Asn Leu Thr Gln Val Leu Lys
      50           55           60
Asp Ala Glu Ser Asp Glu Val Leu Gly Lys Met Lys Val Ser Ala Leu
      65           70           75           80
Leu Glu Ala Leu Pro Lys Val Gly Lys Val Gln Ala
      85           90

```

(2) INFORMATION FOR SEQ ID NO:229:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 72 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:229:

```

Thr Val Glu Leu Asp His Pro Leu Val Glu Val Ser Thr Asp Lys Val
 1           5           10           15
Asp Thr Glu Ile Pro Ser Pro Ala Ala Gly Val Leu Thr Lys Ile Ile
      20           25           30
Ala Gln Glu Asp Asp Thr Val Glu Val Gly Gly Gln Leu Ser Val Ile
      35           40           45
His Asp Ala His Asp Ala Gly His Ala Ala Val Pro Ala Pro Gln Lys
      50           55           60
Val Ser Ala Gly Pro Thr Arg His
      65           70           72

```

(2) INFORMATION FOR SEQ ID NO:230:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 110 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:230:

Ala Ala Asp Ile Gly Ala Ala Pro Ala Pro Lys Pro Ala Pro Lys Pro
 1 5 10 15
 Val Pro Glu Pro Ala Pro Thr Pro Lys Ala Glu Pro Ala Pro Ser Pro
 20 25 30
 Pro Ala Ala Gln Pro Ala Gly Ala Ala Glu Gly Ala Pro Tyr Val Thr
 35 40 45
 Pro Leu Val Arg Lys Leu Ala Ser Glu Asn Asn Ile Asp Leu Ala Gly
 50 55 60
 Val Thr Gly Thr Gly Val Gly Gly Arg Ile Arg Lys Gln Asp Val Leu
 65 70 75 80
 Ala Ala Ala Glu Gln Lys Lys Arg Ala Lys Ala Pro Ala Pro Ala Ala
 85 90 95
 Gln Ala Ala Ala Pro Ala Pro Lys Ala Pro Pro Glu Asp Pro Met
 100 105 110
 Pro

(2) INFORMATION FOR SEQ ID NO:231:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 118 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:231:

Val Leu Val Ser Ile Ser Ala Asp Glu Asp Ala Thr Val Pro Val Gly
 1 5 10 15
 Gly Glu Leu Ala Arg Ile Gly Val Ala Ala Asp Ile Gly Ala Ala Pro
 20 25 30
 Ala Pro Lys Pro Ala Pro Lys Pro Val Pro Glu Pro Ala Pro Thr Pro
 35 40 45
 Lys Ala Glu Pro Ala Pro Ser Pro Pro Ala Ala Gln Pro Ala Gly Ala
 50 55 60
 Ala Glu Gly Ala Pro Tyr Val Thr Pro Leu Val Arg Lys Leu Ala Ser
 65 70 75 80
 Ile Asn Asn Ile Asp Leu Ala Arg Val Thr Gly Thr Gly Val Gly Gly
 85 90 95
 Arg Ile Arg Lys Gln Asp Val Leu Ala Ala Ala Glu Gln Lys Lys Arg
 100 105 110
 Ala Lys Ala Pro Ala Pro
 115

(2) INFORMATION FOR SEQ ID NO:231:

SEQUENCE

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:232:

```

Asp Pro Lys Val Gln Ile Gln Gln Ala Ile Glu Glu Ala Gln Arg Thr
 1           5           10           15
His Gln Ala Leu Thr Gln Gln Ala Ala Gln Val Ile Gly Asn Gln Arg
 20           25           30
Gln Leu Glu Met Arg Leu Asn Arg Gln Leu Ala Asp Ile Glu Lys Leu
 35           40           45
Gln Val Asn Val Arg Gln Ala Leu Thr Leu Ala Asp Gln Ala Thr Ala
 50           55           60
Ala Gly Asp Ala Ala Lys Ala Thr Glu Tyr Asn Asn Ala Ala Glu Ala
 65           70           75           80
Phe Ala Ala Gln Leu Val Thr Ala Glu Gln Ser Val Glu Asp Leu Lys
 85           90           95
Thr Leu His Asp Gln Ala Leu Ser Ala Ala Gln Ala Lys Lys Ala
100           105           110
Val Glu Arg Asn Ala Met Val Leu Gln Gln Lys Ile Ala Glu Arg Thr
115           120           125
Lys Leu Leu Ser Gln Leu Glu Gln Ala Lys Met Gln Glu Gln Val Ser
130           135           140
Ala Ser Leu Arg Ser Met Ser Glu Leu Ala Ala Pro Gly Asn Thr Pro
145           150           155           160
Ser Leu Asp Glu Val Arg Asp Lys Ile Glu Arg Arg Tyr Ala Asn Ala
165           170           175
Ile Gly Ser Ala Glu Leu Ala Glu Ser
180           185

```

(2) INFORMATION FOR SEQ ID NO:233:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 71 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(1) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:233:

```

Val Ser Thr Ser Thr Trp Val Asn His Pro Val Arg Asp Arg Val Ile
 1           5           10           15
Gly Gln Arg Trp Thr Cys Ala Asp Arg Arg Ser Ile Glu Glu Ser Thr
 20           25           30
Gln Met Ala Phe Ser Val Gln Met Pro Ala Leu Gly Glu Ser Val Thr
 35           40           45
Gln Gly Thr Val Thr Arg Trp Leu Asp Gln Gln Gly Asp Thr Val Gln
 50           55           60
Leu Asp Gln Ser Val Val Val Val Val Val Val Val Val Val Val Val
 65           70           75           80

```

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 182 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:234:

Glu Val His Leu Pro Val Asp Val Gly Glu Pro Arg Gln Pro Thr Gly
 1 5 10 15
 Ala Val Ala Asp Gln Asp His Arg Ile Thr Pro Val Pro Ala His Lys
 20 25 30
 His Thr Pro Pro Arg Val Cys Gln Asp Trp His Arg Gln Pro Pro His
 35 40 45
 Arg Gly Arg Ala Asp Gln His Leu Gly Leu Asp Ala Arg Leu Cys Ala
 50 55 60
 Ala Ala Cys Asn Val Leu Leu Val Asp Gly Val Gln His Arg Pro Gln
 65 70 75 80
 Arg His Gly Pro Gly Pro Arg Phe Gly Phe Pro Arg Val Val Val Ala
 85 90 95
 Cys Gly Ile Arg Gln Ala Arg Val Gln Val Glu Arg Phe Gly Gly Val
 100 105 110
 Leu Pro Glu Arg Ala His Gly Val Gly Gln Arg Asn Asn Arg Val Ala
 115 120 125
 Thr Asp Arg Leu Thr Asp Arg Met Pro Ile Asp Arg Gly Leu Gly Arg
 130 135 140
 Glu Pro Arg Ser Val Gly Gly Gln Ile Asp Arg Glu Arg Asp Gln Pro
 145 150 155 160
 Gln Arg Ile Pro Ala Gly Cys His Val Thr Pro His Cys Ser Gln Pro
 165 170 175
 Arg Ser Leu His Leu Val
 180

2. INFORMATION FOR SEQ ID NO:235:

(i) SEQUENCE CHARACTERISTICS:

- A LENGTH: 150 amino acids
- B TYPE: amino acid
- C STRANDEDNESS: single
- D TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:235:

Asn Asp Arg Leu Ile Ser Met Arg Asp Gly Gln Ile Val Ala Leu Pro
 1 5 10 15
 Gln Leu Thr Asp Gln Gln Arg Ala Ala Ala Leu Ile Leu Ala Ala Ala
 20 25 30 35 40

Gly Lys Met Lys Val Ser Ala Leu Leu Glu Ala Leu Pro Lys Val Gly
 65 70 75 80
 Lys Val Lys Ala Gln Glu Ile Met Thr Glu Leu Glu Ile Ala Pro His
 85 90 95
 Pro Ala Ala Phe Val Ala Ser Val Thr Val Ser Ala Arg Pro Cys Trp
 100 105 110
 Lys Ser Ser Ala Pro Pro Asn Pro Ala Gly Arg Arg Cys Gly Pro Glu
 115 120 125
 Gly Leu Trp Trp Ala Tyr Pro Arg Ile Arg Gly Arg Ser Gly Leu Thr
 130 135 140
 Gly Pro Ala His Asn Ser Gly Arg Thr Pro Arg Trp Gly Gly Thr Arg
 145 150 155 160

(2) INFORMATION FOR SEQ ID NO:236:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 178 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(12) SEQUENCE DESCRIPTION: SEQ ID NO:236:

Asp Trp His Arg Gln Pro Pro His Arg Gly Arg Ala Asp Gln His Leu
 1 5 10 15
 Gly Leu Asp Ala Arg Leu Cys Ala Ala Cys Asn Val Leu Leu Val
 20 25 30
 Asp Gly Val Gln His Arg Pro Gln Arg His Gly Pro Gly Pro Arg Phe
 35 40 45
 Gly Phe Pro Arg Val Val Val Ala Cys Gly Ile Arg Gln Ala Arg Val
 50 55 60
 Glu Val Glu Arg Phe Gly Gly Val Val Pro Glu Arg Ala His Gly Val
 65 70 75 80
 Gly Gln Arg Asn Asn Arg Val Ala Thr Asp Arg Leu Thr Asp Arg Met
 85 90 95
 Pro Ile Asp Arg Gly Leu Gly Arg Glu Pro Arg Ser Val Gly Gly Gln
 100 105 110
 Ile Asp Arg Glu Arg Asp Gln Pro Gln Arg Ile Phe Ala Gly Lys His
 115 120 125
 Val Thr Pro His Cys Pro Gln Pro Arg Ser Leu His Leu Val Leu Thr
 130 135 140
 Ser Arg Arg His Val Glu Arg Gln Arg His Arg Ala Glu Glu Gln His
 145 150 155 160
 Glu Val His Ala Gly Pro Leu Gly Gly Ala Ser Gln Ser Gln Ala Ala
 165 170 175
 Pro Arg

- (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:237:

```

ATGCCAAGCC GGTGCTGATG CCCGAGCTCG GCGAATCGGT GACCGAGGGG ACCGTCATTC      60
GTGGGCTGAA GAAGATCGGG GATTCGGTTC AGGTTGACGA GCCACTCGTG GAGGTGTCCA      120
CCGACAAGGT GGACACCGAG ATCCCGTCCC CGGTGGCTGG GGTCTTGGTC AGTATCAGCG      180
CCGACGAGGA CGCCACGGTG CCCGTCGGCG GCGAGTTGGC CCGGATCGGT GTCGCTGCCG      240
AGATCGGGCG CGCGCCCGCC CCCAAGCCCC C

```

(2) INFORMATION FOR SEQ ID NO:238:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 89 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:238:

```

Ala Lys Pro Val Leu Met Pro Glu Leu Gly Glu Ser Val Thr Glu Gly
 1           5           10           15
Thr Val Ile Arg Trp Leu Lys Lys Ile Gly Asp Ser Val Gln Val Asp
 20           25           30
Glu Pro Leu Val Glu Val Ser Thr Asp Lys Val Asp Thr Glu Ile Pro
 35           40           45
Ser Pro Val Ala Gly Val Leu Val Ser Ile Ser Ala Asp Glu Asp Ala
 50           55           60
Thr Val Pro Val Gly Gly Glu Leu Ala Arg Ile Gly Val Ala Ala Glu
 65           70           75           80
Ile Gly Ala Ala Pro Ala Pro Lys Pro
 85

```

(2) INFORMATION FOR SEQ ID NO:239:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 111 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:239:

TAATGAGCGG ATGCGCGGGA ATGCGCGGGA ATGCGCGGGA ATGCGCGGGA ATGCGCGGGA

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 339 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:240:

ATGAAGTTGA AGTTTGCTCG CCTGAGTACT GCGATACTGG GTTGTGCAGC GCGCCTTGTG 60
 TTTCCTGCCT CGGTTGCCAG CGCAGATCCA CCTGACCCGC ATCAGCCGGA CATGACGAAA 120
 GGCTATTGCC CGGGTGGCCG ATGGGGTTTT GCGCACTTGG CCGTGTGCGA CCGCGAGAAG 180
 TACCCCGACG SCTCGTTTTG GCACCACTGG ATGCAAACGT GGTTTACCGG CCCACAGTTT 240
 TACTTCGATT GTGTCAGCGG CCGTGAGCCC CTCGCCGGCC CGCCGCCACC GGGTGGTTGC 300
 GGTGGGGCAA TTCCGTCCGA GCAGCCCAAC GCTCCCTGA 339

(i) INFORMATION FOR SEQ ID NO:241:

(ii) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 112 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(iii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:241:

Met Lys Leu Lys Phe Ala Arg Leu Ser Thr Ala Ile Leu Gly Cys Ala
 1 5 10 15
 Ala Ala Leu Val Phe Pro Ala Ser Val Ala Ser Ala Asp Pro Pro Asp
 20 25 30
 Pro His Gln Pro Asp Met Thr Lys Gly Tyr Cys Pro Gly Gly Arg Trp
 35 40 45
 Gly Phe Gly Asp Leu Ala Val Cys Asp Gly Glu Lys Tyr Pro Asp Gly
 50 55 60
 Ser Phe Trp His Gln Trp Met Gln Thr Trp Phe Thr Gly Pro Gln Phe
 65 70 75 80
 Tyr Phe Asp Cys Val Ser His Gly Gln Pro Leu Pro Gly Pro Pro Pro
 85 90 95
 Leu Gly Gly Tyr Gly Gly Ala Ile Pro Ser Gln Gln Pro Asp Ala Pro
 100 105 110 112

(i) INFORMATION FOR SEQ ID NO:242:

(ii) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 371 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

```

GTGACCACGG TGGGCGTGGC ACCAACCCGG GCAGCGGCAG CCGCGGCGGG GCGGCGGGCT 60
CCGGCGGCCAA CGGTGGGCGC GGGGGTAACG CCACCGGCTC AGGCGGCAAG GGCGGCGCG 120
GTGGCAATGG CGGTGATGGG AGCTTCGGCG CTACCGAGCG CCGCGCTCC ATCGGGGTCA 180
CGGGCGCCCC CGGCGGCAAC GGCGGCAAGG GCGGCGCCCG TGGCAGCAAC CCGAACGGCT 240
CAGGTGGCGA CGGCGGCAAA GGCGGCAACG GCGGTGCCCG CGGCAACGGG GGCTCGATCG 300
GCGCCAACAG CGGCATCGTC GCGGTTCG GTGGGGCCCG TGGCGCTGGC GGCGCCGCG 360
GAAACGGCAG C

```

(2) INFORMATION FOR SEQ ID NO:243:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 424 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:243:

```

GTCCGGGTCC CACCACCCCG CCGGCGCGGC CCTAGCGGCC GGGCGCACCA GCCCCTTTTC 60
TTGACTCGTT CAAGAAAAGG GCCTTCTGTT TGGTCGGCCA TGTTGGCATG ATCGTGACCC 120
ATGGGCAACA TCGACGTGCA CATCTCGGCC AAGGTCTAGC TCCATGCGAA TCGCCGCGCG 180
GGTGGTGAGC ATCGGTCTAG CCGTCATAGC AGGGTTCGCG GTACCTGTTG CCGACGCACA 240
CCCGTCGGAG CCGGGGGTTG TGTCTACG GGTGCTCGGA AAGGGGTGCG TCGGCAACAT 300
CGTCGGCGCC CCAATGGGCT GGGAGGCGGT GTTCACCAAG CGTTTCCAGG CGTTTGGGT 360
CGAACTACCG GCGTCACCA ACTGGGTGGA CATCGGGCTG CCGAGGTGT ACGACGATCC 420
CGAC

```

(2) INFORMATION FOR SEQ ID NO:244:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 317 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:244:

```

TCGATGGGCG CCGCGGGTAA CAGGCGCAAT TTGGAACGGT TTGCGAAGCG CAAAGATGCT 60
TTGATCTGCG CCGCAATTGA TTGAGCCCGG GCGGATTTCG TCACCGAGGG CCACCGTCTA 120
AGGGCGGATG CGATCCTACT GCGCGGTACC GACCGGCTGC GTTTCGCGGA GCGCGCGGAT 180
TGGGATTGCG TCGAGTCCCA TTGCGGCACG ACCGTGACCG CCGACACGGT GCGCATCGAC 240
GTGATGCGCG ACGATATGCG TTGCGAACTG GCGGCGGCGT CCAAACTCA CCAATCGCTG 300
CGGCTCTAGC ATTGCTC

```

(2) INFORMATION FOR SEQ ID NO:245:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:245:

TGGCGTATGC	GCTTCGCAGC	CGGTGCCGCG	TCAACGCGCC	GGAGGCAATC	GCTTCGCTGC	60
CGAGGAATGG	TTGATCAGC	ATCGCAGTGT	GCCGTCTGTC	ACCGACACCG	CCGTCCAACG	120
TGAACTGAGG	GCGGAAAATC	GGCCGAAATC	TGCGCCTCAG	TTACAGCTCG	GCGCCTAACG	180
GTTCCTGGAAG	TTGGGTGCGC	GCTTCTCGGC	GAACGCGCGC	GGGCCTTCCT	TGGCGTCTGC	240
GGACAGGAAG	ACCTTGATGC	CGATCTGGGT	GTCCATCTTG	AACGCCTCGT	TTTCGGGCAT	300
GCACTCGGTC	TGCGGATGG	ACCGCAAGAT	GGCCTGCACG	GCCAGGGGTC	CGTTAGCCGA	360
GATGGCGTCG	GCAAGTTCTA	GAACCTTGGT	CAACGCCTGG	CCGTGCGGCA	CACGTGGCCG	420
AT						422

(2) INFORMATION FOR SEQ ID NO:246:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 426 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:246:

GGTGGCGCT	GAACACCAGC	CGCGGGCTGC	CAGATCTCCC	GGACTCGGTA	GTGCGGCGGG	60
TGGCGTCTGT	GCTCTCTGA	CGGGGCGCGG	CGACCATAAG	GTGCTAATG	CCCAGGTAGC	120
GGCCAGGTG	CATGGAGTGC	ATGATGATGC	GACTCTCCAG	CTGCGCGACC	GGGAGCTTGG	180
CATCGGGGCT	GATCAGCCAG	GACCGGTAGG	ACAAGTGGAT	CGAATGCATA	GTGGCCTCCA	240
GAGTGGCGCT	GGCACTTCCG	CGGTGCTCCA	CGGCAAATGC	CTTGATTCT	AGCTCCGCGT	300
AGTGTTCGGG	CATCGGCTGC	GGGATGAATG	GGAACCGCAG	GATGGCGACA	AACGGGTCTG	360
ACCTGAGGT	TGCGGCTTTC	GGCAGCTGG	TGCACAGCCG	GTACTCGGCA	TAAATGCTGG	420
CCCCGA						426

(2) INFORMATION FOR SEQ ID NO:247:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 327 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:247:

AGACCGGCCA	GGGTGTGGTC	GCTGCGGCGG	GGATTCTGGA	TAATCTGCCG	TGGGTGAGC	60
GGCGGATGAA	CTAGTGAGGT	GCACGCTTAC	GGTTTGGGAT	AGCCACAGCT	AAAAAGTTTA	120
TCAAAGAAAC	GAAGAAGGTT	TGATGAGTAA	CTTTTGGGCT	TACCGCGGCG	TGCTCGGCGG	180

(2) INFORMATION FOR SEQ ID NO:248:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 123 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:248:

```

Asp His Gly Gly Pro Ala Thr Asn Pro Gly Ser Gly Ser Arg Gly Gly
 1           5           10           15
Ala Gly Gly Ser Gly Gly Asn Gly Gly Ala Gly Gly Asn Ala Thr Gly
 20           25           30
Ser Gly Gly Lys Gly Glu Ala Gly Gly Asn Gly Gly Asp Gly Ser Phe
 35           40           45
Gly Ala Thr Ser Gly Pro Ala Ser Ile Gly Val Thr Gly Ala Pro Gly
 50           55           60
Gly Asn Gly Gly Lys Gly Gly Ala Gly Gly Ser Asn Pro Asn Gly Ser
 65           70           75           80
Gly Gly Asp Gly Gly Lys Gly Gly Asn Gly Gly Ala Gly Gly Asn Gly
 85           90           95
Gly Ser Ile Gly Ala Asn Ser Gly Ile Val Gly Gly Ser Gly Gly Ala
100           105           110
Gly Gly Ala Gly Gly Ala Gly Gly Asn Gly Ser
115           120

```

(2) INFORMATION FOR SEQ ID NO:249:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 104 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:249:

```

Met Ala Ala Ala Gly Thr Thr Ala Asn Val Val Arg Phe Pro Asn Pro
 1           5           10           15
Asn Asp Pro Leu His Leu Ala Ser Ile Asp Phe Ser Pro Ala Asp Phe
 20           25           30
Val Thr Glu Gly His Arg Leu Arg Ala Asp Ala Ile Leu Leu Arg Arg
 35           40           45
Thr Asp Arg Leu Pro Phe Ala Gly Pro Pro Asp Thr Asp Leu Val Glu
 50           55           60
Ser Glu Leu Arg Phe Thr Val Thr Ala Asp Thr Val Arg Ile Asp Val
 65           70           75

```

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:250:

Ala Tyr Ala Leu Arg Ser Arg Cys Arg Val Asn Ala Pro Glu Ala Ile
1 5 10 15
Ala Ser Leu Pro Arg Asn Gly Ser Ile Thr Ile Ala Val Cys Arg Arg
20 25 30
Ala Pro Thr Pro Pro Ser Asn Val Asn
35 40

(2) INFORMATION FOR SEQ ID NO.251:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:251:

Val Pro Leu Asn Thr Ser Pro Arg Leu Pro Asp Leu Pro Asp Ser Val
1 5 10 15
Val Pro Pro Val Ala Ser Leu Leu Ser
20 25

2) INFORMATION FOR SEQ ID NO:252:

2. SEQUENCE CHARACTERIZATION

- A LENGTH: 61 amino acids
B TYPE: amino acid
C STRANDEDNESS: single
D TOPOLOGY: linear

... MOLECULE TYPE: protein

(XX) SEQUENCE DESCRIPTION. SEQ ID NO: 252.

Met Ser Thr Val Ala Ala Tyr Ala Ala Met Ser Ala Thr Glu Pro Leu

Gln Thr Glu Trp Gly Gln Pro Asn Leu Pro Val Val Pro
 50 55 60

(2) INFORMATION FOR SEQ ID NO:253:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 213 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:253:

```

GCTTGGAGCC CTGGAGCGAC GGTGTGGGTC TGGGGGTGGA TTCGTTCTCG GCGAAAGTCA    60
ACTAAAGACC ACGTTGACAC CCAACCGGCG GCCCGGCATG GCGCGTCGCG GCGTAGAAGC    120
TTTGACCGCG GCGCGAAACG TTCGCTGCTG CGGCCCATGC AGATCGCACA CGCTTGCTTG    180
AACATCGGGT GGAGCCGGTG CTAACGCCAG GCT                                213

```

(2) INFORMATION FOR SEQ ID NO:254:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 367 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:254:

```

CGGAGCTGCT GTTCGGCGCC GCGCGTGGCG GCGCGCGCGG TGGGGGCGGC ACCGACGGCG    60
GGCCCGGTGC TACCGCGCGG ACCCGCGGAT ACGGCGGAGT CGCGCGCGAC GCGGATGGC    120
TGGCACCCCG CGGGGCGCGC GGGGCGCGCG GCGAAGCGCG GCGAGTTGGT GCGCGCAGCG    180
ATGGTGGCGC GTTGGGTGGT ACCGCGCGGA TGGCGGTAC CGGCGCGCGC GGTGGCGCGC    240
GCGGTGCGCG CACACTGCTG CTGGGCGCTG GCGGACAGGG CGGCTTCGGC GCGCGCGCGC    300
GACAAAGGCG CACCGCGCGG GCGCGCGCGA GATGGCGTTG TGGGGGTTGT CAGTGGCACT    360
GTGGTA

```

(2) INFORMATION FOR SEQ ID NO:255:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 420 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

```

GTCCGCTCTA TGGGACGGGT GCGACGCGCG ATTGTCCAAC CGCCTGGGCT ACCCCGAAGT 240
CCGCGCGCGCA CTCGCTGCAA CGGGCGCGCAA TCACGACCTA ACCGAATCCG AGCTCGCGCA 300
CGCCGAGCGT GACTGGGAGG ACTTCTGGGC CGCACCCGCG CAGTCGAACT CACCGCGACG 360
GTTGAACAGC ACGCCGGGCA CCTCGCCCGA ACACATGCCT TACGCGGAGC CGACACCGTT 420

```

(2) INFORMATION FOR SEQ ID NO:256:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 299 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:256:

```

CTTTTGTCCG TGGCATCCGC GGTACCGGCG GAACCGGGCG CAACCGCGGT ATGCTCGCGG 60
GGCGCGCGCG GCGCGCGCGT GCGCGCGGGT TCAGCTTCAG CACTGCGCGT GGGGCTGGCG 120
CGCGCGCGCG GCGCGGTGGG CTGTTACCCA CCGCGGGTGT CGCGCGCGCG GGTGGGCAGG 180
GTCACACGGG CGGGCGCGCG GCGCGCGGCG GGGCGCGCGG GTTGTTCGT GCCGCGCGCA 240
TGGCGCGGGG GGGCGGATTC GGGGATCAGG GAATGCTCGG CACCGCGCGG GCGCGCGGG 299

```

(2) INFORMATION FOR SEQ ID NO:257:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:257:

```

Leu Glu Pro Trp Ser Asp Gly Val Gly Leu Gly Val Asp Ser Phe Ser
1           5           10           15
Ala Lys Val Asn
20

```

(2) INFORMATION FOR SEQ ID NO:258:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 121 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:258:


```

      20      25      30
Val Gly Gly Asp Gly Gly Trp Leu Ala Pro Gly Gly Ala Gly Gly Ala
      35      40      45
Gly Gly Gln Gly Gly Ala Gly Gly Ala Arg Ser Asp Gly Gly Ala Leu
      50      55      60
Gly Gly Thr Gly Gly Thr Gly Gly Thr Gly Gly Ala Gly Gly Ala Gly
      65      70      75      80
Gly Arg Gly Thr Leu Leu Gly Ala Gly Gly Gln Gly Gly Leu Gly
      85      90      95
Gly Ala Gly Gly Gln Gly Gly Thr Gly Gly Arg Arg Arg Trp Arg
      100      105      110
Ser Gly Gly Cys Gln Trp His Trp Trp
      115      120

```

(2) INFORMATION FOR SEQ ID NO:259:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:259:

```

Gly Val Ile Gly Lys Ala Thr Ala Gln Arg Pro Val Ala Ala Gly Arg
 1           5           10           15
Pro Arg Pro Arg Pro Gln Arg Pro Val Ser Asp Arg Val Ser Asp Gln
      20           25           30
Arg Arg

```

2. INFORMATION FOR SEQ ID NO:260:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 99 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:260:

```

Leu Val Gly Gly Ile Gly Gly Thr Gly Gly Thr Gly Gly Asn Ala Gly
 1           5           10           15
Met Leu Ala Gly Ala Ala Gly Ala Gly Gly Ala Gly Gly Phe Ser Phe
      20           25           30
Ser Thr Ala Gly Gly Ala Gly Gly Ala Gly Gly Ala Gly Gly Leu Phe
      35           40           45

```

Gly Gly Ala Gly Gly Phe Gly Asp His Gly Thr Leu Gly Thr Gly Gly
 85 90 95
 Ala Gly Gly

(2) INFORMATION FOR SEQ ID NO:261:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 282 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:261:

TCCTGTTCCG	CGCCGGCCGG	GTGGGCGGTG	TTGGCGGTGA	CGGTGTGGCA	TTCTGGGCA	60
CGGCCCCCGG	CGGGCCCCGT	GGTGCCGGCG	GGGCGCGTGG	GCTGTTGAGC	GTCCGGTGGGG	120
CCGGCGGGCG	CGCGCGGAATC	GGATTGGTCC	GGAACAGCGG	TGCCGGGGGG	TCCGGCGGGT	180
CGGCGCTGCT	CTGGGGCGAC	GGCGGTGCCG	GCGGCGCGGG	TGGGCTCGGG	TCCACTACCG	240
GGGGTSCCGG	CGGGGCGGGG	GGCAACGCCA	GCTGCTGGT	AA		282

(2) INFORMATION FOR SEQ ID NO:262:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 415 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:262:

CGGCACGAGC	CGTGCTACTG	GTCAACTGAT	CGCTGATTG	TGACCTTCCG	GGCGCCGGAT	60
CAGTGCTTCT	CAGGACCGAC	GTAAATATTC	AAACCAATC	CGGCGCCCGA	GGCGAGGATG	120
AATGCCACAC	CGGCGGCGAT	CAGCCACGGG	AGCCACAACG	CGATGCCGAC	CGCTGCCACC	180
CAGCCGGACA	ACGCGACCAT	GATCGGCCAC	CAGCTATGCC	GACTGAAGAA	TCCAAATTCT	240
CGTGGCGCGT	CGCTGATTTC	AGGCGCTTCT	TAGTCCCTCG	CGCGGGAATC	TAACCGGCGG	300
AGACAAACG	CGAAGAAGGT	GGCGACGATC	AACGCCATGC	CGCGGCTGAG	CGCCCAACGCA	360
ATGGTCCGAG	CGCACTCGAC	AGCACTGCTG	CGAAGATCG	AGGTCAACAC	CGCT	415

(2) INFORMATION FOR SEQ ID NO:263:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 373 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

TCACCGCGTG AACGGTTCGT AACACTGATA CGTATGCTTG TCAGCGAGCA GATCAAGTCC	60
AGTCCGACCA ATGCCAGGAG ATCATCGGCT AGGCTCACGG TTTCGCCTGG GACGAGACGG	120
TATTGAGTTC TGGCSTTGA CGGTCCGTGG CGTGGTGGGA AGTCTGACGC GGCATCAGAA	180
CGGTGTGCAA TACCAGTCTT TGGGGGATAT GGCCTATTTG GTGTCGTGG GCCGCTCCAC	240
CGGATCCCTT TTCGAACGTT GC3CAAGCGC GGTCCAGTTA CGGCCTGTT ACTGCGCGCT	300
GGCGTAGCTG CGCGGCTCG ATCGGTTTGA ACGTCATCGC AATTCGCCGA ATGGGTGAGT	360
ACCTGACGCT CCT	373

(2) INFORMATION FOR SEQ ID NO:264:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 423 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:264:

GCAAACCGGA CAGGCCGGCA GCGACGGTCG GAAGTTGCGC CAUGGTGCGC GCTCCATGTA	60
GCAAACCGGT GACCACGGCG TAGACAGCAG ATCCGTGGAT CGGCGGTTCC GTGTCGTCCG	120
GGCCGAGTAC CCGCGGGCCG AACCGCAGCG ACCAAAGCAA CGCGATCGAT ACGGGGATCG	180
CCACTCGTGC CCAATTGAG CTCCCTCGAC AAGCTTGCGG CCGCACTCGA ACCCGGGTGA	240
ATGATTGAGT TTAACCGCT TAGCAATAAC TAGCATAACC CCGTGGGGCC TCTAAACGGG	300
TCTTGAGGGG TTTTCTGCTG AAAGGAGGAA CTATATCCGG ATAACCTGGC GTAGTAGCGA	360
AGAGGCCCGC ACCGATCGCC CTTCGCAACA GTTGCGGAGC GTGAATGGCG AATGGACGCG	420
CCC	423

(2) INFORMATION FOR SEQ ID NO:265

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 404 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:265:

ACTGSCCAGC CGGTGCGCA ATCCATCCAG TTTCTTTTAT CTCAGTGCAG CATCCGCGCA	60
ACTGACCGCG ACCGAGTCCG CTCTCTTCGGC AGGCAATTTG TCGAAGCCGG TATGCAACCG	120
CGGTCCCGAC GTGCTACAT CCGGCAGCG AGGTGCGATG GCATGCTGCT CGCCGTCCAG	180
CGAATGTGCG ACCTGCGGCA CCGCCCGGATG CGACCGGCTG ACCAAGCCGT GTAACACAGC	240
CAGCACCGCG CTGCGGAGGC TTTCCGGCGG CATGCTGCG ACCGCACCGT CGAGCAGCTC	300
CAGTACGAGC GTGAGGTGAC CGGTGCTGCG CTCGCGCGCG AGGCTCAGCG GAAAGTGCGA	360
CGAAGTCTCT AGCGCGACCG GAGCGAACGT CAGCGCTTTT GCGA	404

(2) INFORMATION FOR SEQ ID NO:266

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:266:

GTCTCTGGTCC	CAGGCTGTTC	TTGGAACCCG	CTGGCTAACT	TCCACCCCGG	GTATCCGCCC	60
ACCATCGAAC	CCGCCCAACC	GGCGGTGTCA	CCGCCTACTT	CGCAAGACCC	GGCCGGTGCA	120
GTGCGACCAC	TGAGCGGCCA	CCCCCGGGCG	GCACTATTCC	ACAACGGCAC	CCGCCAATTG	180
GTGGCTCTGC	GCCCGGGCGC	CGATTGGGCG	GCACCCGCCA	GCATCATGGT	CTTCGATGAC	240
ATGCACGTTG	CACCGCGCGT	CATTTTCTG	CCGGGCCCGG	CAGCCGCGTT	GACCAGCGAC	300
GACCACGGCA	CGGCCTTCCT	TGCCGCCCGC	GGCGGCTACT	TCGTGGCCGA	CCTGTCTCTC	360
GGTCACACCG	CACGAGTGAA	TGTCGCTGAC	GCAGCGCACA	CCGATTTTAC	CGCGATCGCC	420
C						421

(2) INFORMATION FOR SEQ ID NO:267:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 426 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:267:

ATGCATATCA	CGCTCAACCG	CATCTGCCCT	GCGATCTTCG	GGGCCGGCGG	CAGTGAACCTA	60
GACGAGCTGC	GCCGCCCTCAT	TCCGCCGTGG	GTACCGCTGG	GCTCGCGCCT	GGCGGCGCTA	120
CCGAAACCCA	AACGCGACTA	TGGCGGCCTT	AGCCCTGGG	GCCGGCTGGC	CGAGTGGCGG	180
CCCGAGTACG	ACACTGTCTA	CGACGAGCTC	ATCGAAGCCG	AGCGGGCCGA	CCCGAACTTC	240
CCCGATCGGA	CCGACSTTTT	GGCGTTGATC	CTGCGCAGCA	CTTACGACGA	CGGTTCCATC	300
ATCTCGCGCA	AGGACATTGG	CGACGAACTC	CTCAGGCTGC	TTGCCGCCCG	GCACGAAACC	360
ACCGCGGCGA	CATGGGCTGG	TCCTTCGAAT	CGCTCAACCG	GCACCCCGAA	GTGCTCGCGG	420
CTCTGG						426

(2) INFORMATION FOR SEQ ID NO:268:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 426 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:268:

GTCTCTGGTCC	CAGGCTGTTC	TTGGAACCCG	CTGGCTAACT	TCCACCCCGG	GTATCCGCCC	60
ACCATCGAAC	CCGCCCAACC	GGCGGTGTCA	CCGCCTACTT	CGCAAGACCC	GGCCGGTGCA	120
GTGCGACCAC	TGAGCGGCCA	CCCCCGGGCG	GCACTATTCC	ACAACGGCAC	CCGCCAATTG	180

CGCCGCTCCG ACGGCAAGCT GGTGCTGGGC AGCGCAGATG GCGCCGTCTA CACGCTTGCC 480
AAGAACCCGC AGTTGACCGG CGTCGGCGGC GCCACCGTAG CC 522

(2) INFORMATION FOR SEQ ID NO:269

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 739 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:269:

GGTGGGGCGC ACCGCCGTCC GGGGGCCCCA GCGCCTGGGC CCAGACCCCG GCGAAAACCA 60
 ACCCGTGGCC CTACTGGCC GGGGCGCCG CCGTCGTGCT CGTCCTCGTG TTGGGCGCCA 120
 TCGGCATCTG GATCGCCATC CGGCCCAAGC CGGTACAGCC GCCTCAGCCG GTTGCGGAGG 180
 AGCGCTTTAG CGCCCTACTG CTGAAGTCCT CAGAAGTCAA GCGCCGTGATG GGCTCGTCTG 240
 CCATGCAAGC GGGCAAACCG ATCAGATCGA TGGACTCTTC GCGCGTGACG GTGTCCCTGC 300
 CGGACTGCCA GGGCGCGCTG TATACCAGCC AGGATCCGGT GTATGCCGGC ACCGGGTACA 360
 CCGGCATCAA CGGCTTGATT TCATCCGAGC CGGGCGACAA CTACGAACAT TGGGTGAACC 420
 AAGCGTCTGT CGCCTTTCCG ACCGCCGACA AAGCCCGCGC GTTCGTGCAG ACTTCGGCCG 480
 ACAATGGAA GAATGCGCA GGCAGACCG TCACCGTCAC GAATAAGGCC AAGACCTACC 540
 GGTGGACGTT TGCCGACGTC AAAGGCAGCC CGCCGACGAT CACGGTGATA GACACCCAAG 600
 AAGGCGCTGA GGGCTGGGAA TGCCAACGCG CGATGAGCGT GGCCAACAAT GTGGTTGTCT 660
 ACSTCAACGC ATGCGGGTAC CAGATCAGCA ATCAAGCAGG CCAGATCCGC GCCAAGATCT 720
 GTTGACAAAG TCAACAAGC 739

(2) INFORMATION FOR SEQ ID NO:270:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 69 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:270:

AAGATCTCTG GAGGCGCGCA GCGGCGCGC TGAATCTCTA AAGGCGCGCA TGAACCTCTT 60
 GGTATCTG 69

(2) INFORMATION FOR SEQ ID NO:271:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 523 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

ACTGCACCCG	GCAGGCGGCA	CCAACGGATG	GGGTCAACTA	GCACTGCCCG	TGGAGGCGCC	60
CCCGCGGTCT	GTGCCCTTCC	ACGGGGAACG	TTTGGGCAGC	GCGGCTCCAG	AAGGGTTGGA	120
GGGAGAGTTC	GACGACCGTA	TCGACGAGCG	GTTCCCGGTC	TTGAGCTCCG	CCAGTCTCGC	180
CGAAGCGCTG	CCGGGTCCGC	TGACCCCGAT	GACGCTGGAT	GTCCAGTTGA	GTGGACTGCG	240
CGCGGCCGGT	CGGGCGATGG	GTCGGGTACT	GGCGCTTGGC	GGTGTGTTG	CCGATGAGTG	300
GGAGAGAAGA	GCCATCGCGG	TGTTCCGTCA	CCGCCCCTAT	ATCGGAGTGT	CGGCCAATAT	360
TGTGGCCGCC	GCCCAACTGC	CGGGGTGGGA	CGCGCAGGCC	GTAACCCGGC	GGGCACTGGG	420
CGAGCAACCG	CAGGTCATG	AGCTGCTTCC	GTTTGGTCTG	CCGCAACTTG	CGGGCGGACC	480
GCTCGGCTCG	GTCGCAAGG	TGGTCTGAC	GGCACGGTCG	CTG		523

(2) INFORMATION FOR SEQ ID NO:272:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 224 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:272:

CTCTCGGTGT	CGTCGGGGTA	GGAGCGACTT	CCCGGGCCGG	CGCCGGCGCC	GGAGCGGGCT	60
CTGCAGGAAC	CGGTCCCGGC	GTCGCGCGCG	GGGCGACCAA	AGGCCGGATC	GATTCGGCCA	120
CGGCGCTTGG	CGCGCGCTTG	TCCAGCGGCT	TGTTGGCGGT	CCCGAGCCAT	ACCACAAACC	180
AACGCTGAAG	GGGCGCGCGC	TCCGCTGCTT	TCCCGCGCGG	CGAC		224

(2) INFORMATION FOR SEQ ID NO:273:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 521 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:273:

TTAACTGACT	TCCTCTCTTG	ATGCGCGCGC	TTCTTCTTTC	ATAGCTCGCC	TCCTAGGCTA	60
TTAACTGCTC	TTCTCTATAG	CTCTCTCTTG	TTCTTCTTTC	CTCTAAAGCC	TCCTAGGCTA	120
TTAACTGCTC	TTCTCTATAG	CTCTCTCTTG	TTCTTCTTTC	CTCTAAAGCC	TCCTAGGCTA	180
TTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	240
TTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	300
TTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	360
TTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	420
TTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	480
TTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	CTCTCTCTTG	521

(2) INFORMATION FOR SEQ ID NO:274:

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:274:

CTCAGGCTC	ATTCGCTCGA	ACAAAGCCAC	CCGCCCGTAC	AGCGGACGCC	CCCATTCTGTT	60
GTCGTGATAG	TCGCGGTACA	GCTGGGCATC	GGGCCCTGGA	CGAACCTCCG	CCCAGGGGCA	120
GCGAACCAGC	CCGTGCGCCG	TCACGCGGGG	TCAGAACGGT	AGTGCACGAC	AGTCTCGCCG	180
CCCGAAGGGT	TTGACGCGTC	AGACTCGGCC	TCGGCGTCTT	CCGACGAGGC	GTGGATCGCC	240
CCGAGCTGAG	AGCGTAGCCG	CTCGAGCTCA	CGGCCGAGCC	GTTCACGAC	CCAGTCCACC	300
TCGCTGGTCT	TGTTCCCGCG	CAGCACCTGC	GTGAACCTGA	CCGCGTCCGAC	ATCGGCSCGG	360
GTGACCCCGA	ACGCCGGCAG	CSTCSTCGCC	GTCTCSCCC	GCGGCAGGGG	CGGCAACTGC	420
TCGCCA						426

(x) INFORMATION FOR SEQ ID NO:275:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 219 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:275:

GCGGACACGG	CGGACAAAGC	GCAATGSSCC	TCGGCGGCGG	CGCGGGCGGC	GACGGGGGGC	60
AGGGCGGGCC	CGGCCGCGGA	CTGTGGGJTA	CTGGCGGCGC	CGGCGGACAC	GGCGGGGCAA	120
CGCGGTGGTA	CGGGGGGCCC	ACCGCTGCCC	GCTCAGGCAG	GCATGGGCGC	CGCGGTGGC	180
GCGGTGGGT	TCATCGGCAA	CGCGGGGCCC	GCGGGCGAC			219

(x) INFORMATION FOR SEQ ID NO:276:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 571 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:276:

AAGATCATCG	GCCTGGCTGC	TTAGCATCTT	TGAGCTCTCT	ATGCTGCGCG	GCGCGGATCA	60
TCGAGGCTCG	TCCTTCTAGC	CCACTGCTCG	AAAGGTCAGC	ACCACAGTGG	GCTTCTCGGG	120
ATCGTTTCTG	ACCTTGGCCG	GCAGACGGTG	TATATGCGCC	TTGACGAGCC	TGGTATCGCG	180
TGGGTGCGCG	TAACTGCATA	CTCTTTGTA	TAGTACATCA	CGAGTAAACA	GCTGGGCGCG	240
CTTCCGCGCG	AATTTAGACA	ACAGCTCGAA	TTTCAGCGCT	CTCAACGAGA	TCTGCTCAGC	300
ATTCGAGTCT	AGCTTCTGTA	CGCTTCTGTA	CTCTTCTGTA	CTCTTCTGTA	CTCTTCTGTA	360

(2) INFORMATION FOR SEQ ID NO:277:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 93 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:277:

```

Leu Phe Gly Ala Gly Gly Val Gly Gly Val Gly Gly Asp Gly Val Ala
 1             5             10             15
Phe Leu Gly Thr Ala Pro Gly Gly Pro Gly Gly Ala Gly Gly Ala Gly
 20             25             30
Gly Leu Phe Ser Val Gly Gly Ala Gly Gly Ala Gly Gly Ile Gly Leu
 35             40             45
Val Gly Asn Ser Gly Ala Gly Gly Ser Gly Gly Ser Ala Leu Leu Trp
 50             55             60
Gly Asp Gly Gly Ala Gly Gly Ala Gly Gly Val Gly Ser Thr Thr Gly
 65             70             75             80
Gly Ala Gly Gly Ala Gly Gly Asn Ala Ser Leu Leu Val
 85             90

```

(2) INFORMATION FOR SEQ ID NO:278:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:278:

```

Met Pro Pro Val Ser Ala Asn Ala Met Val Glu Val His Ser Thr Pro
 1             10
Met Val Ala Asn Ile Glu Val Asn Thr Pro
 20             26

```

(2) INFORMATION FOR SEQ ID NO:279:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

Lys Pro Asp Arg Pro Ala Ala Thr Val Gly Ser Cys Thr Thr Val Arg
 1 5 10 15
 Ala Pro Cys Ser Gln Pro Val Thr Thr Ala
 20 25

(2) INFORMATION FOR SEQ ID NO:280:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:280:

Trp Pro Ala Gly Arg Pro Met His Pro Ala Pro Gly Thr Ser Ala Asp
 1 5 10 15
 His Pro Pro Asn
 20

(2) INFORMATION FOR SEQ ID NO:281:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 140 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:281:

Val Leu Val Ala Gly Asp Ser Ser Asn Pro Leu Ala Asn Phe Ala Pro
 1 5 10 15
 Gly Tyr Pro Pro Thr Ile Gln Pro Ala Gln Pro Ala Val Ser Pro Pro
 20 25 30
 Thr Ser Gln Asp Pro Ala Gly Ala Val Arg Pro Leu Ser Gly His Pro
 35 40 45
 Arg Ala Ala Leu Ser Asp Asn Val Thr Arg Gln Leu Val Ala Leu Arg
 50 55 60
 Trp Gly Ala Asp Ser Ala Pro Ala Ser Ile Met Val Phe Asp Asp
 65 70 75 80
 Met His Val Ala Pro Arg Val Ile Phe Leu Pro Gly Pro Ala Ala Ala
 85 90 95
 Leu Thr Ser Asp Asn His Gly Thr Ala Phe Leu Ala Ala Arg Gly Gly
 100 105 110
 Tyr Phe Val Ala Asp Leu Ser Ser Gln His Thr Ala Arg Val Asn Val
 115 120 125
 Ala Asp Ala Ala His Thr Gln Thr Ser Thr Thr Thr Thr Thr Thr Thr

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 142 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:282:

```

Met His Ile Thr Leu Asn Ala Ile Leu Arg Ala Ile Phe Gly Ala Gly
 1           5           10           15
Gly Ser Glu Leu Asp Glu Leu Arg Arg Leu Ile Pro Pro Trp Val Thr
 20           25           30
Leu Gly Ser Arg Leu Ala Ala Leu Pro Lys Pro Lys Arg Asp Tyr Gly
 35           40           45
Arg Leu Ser Pro Trp Gly Arg Leu Ala Glu Trp Arg Arg Gln Tyr Asp
 50           55           60
Thr Val Ile Asp Glu Leu Ile Glu Ala Glu Arg Ala Asp Pro Asn Phe
 65           70           75           80
Ala Asp Arg Thr Asp Val Leu Ala Leu Met Leu Arg Ser Thr Tyr Asp
 85           90           95
Asp Gly Ser Ile Met Ser Arg Lys Asp Ile Gly Asp Glu Leu Leu Thr
100          105          110
Leu Leu Ala Ala Gly His Glu Thr Thr Ala Ala Thr Trp Ala Gly Arg
115          120          125
Ser Asn Gly Ser Thr Gly Thr Pro Thr Cys Ser Arg Leu Trp
130          135          140

```

(x) INFORMATION FOR SEQ ID NO:283:

1. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 163 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:283:

```

Val Leu Val Ala His Val Ser Ser Asn Trp Leu Ala Asn Pro Ala Ser
 1           5           10           15
His Tyr Pro Pro Thr Ile Glu Trp Ala Glu Trp Ala Val Ser Pro Pro
 20           25           30
Thr Ser Glu Asp Pro Ala Gly Ala Val Arg Pro Leu Ser Gly His Pro
 35           40           45
Arg Ala Ala Leu Phe Asp Asn Gly Thr Arg Glu Leu Val Ala Leu Arg
 50           55           60
Phe Gly Ala Arg Ser Ala Ala Thr Ala Ser Ile Met Val Phe Asn Asn
 65           70           75           80

```

```
Tyr Phe Val Ala Asp Leu Ser Ser Gly His Thr Ala Arg Val Asn Val  
      115                      120                        125  
Ala Asp Ala Ala His Thr Asp Phe Thr Ala Ile Ala Arg Arg Ser Asp  
      130                      135                        140  
Gly Lys Leu Val Leu Gly Ser Ala Asp Gly Ala Val Tyr Thr Leu Ala  
      145                      150                        155                        160  
Lys Asn Pro
```

(2) INFORMATION FOR SEQ ID NO:284:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 240 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:284:

[illegible]

238

- (A) LENGTH: 22 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:285:

```

Asp Val Val Glu Ala Ala Ile Ala Arg Ala Glu Ala Val Asn Pro Ala
 1             5             10             15
Leu Asn Ala Leu Ala Tyr
                20

```

(2) INFORMATION FOR SEQ ID NO:286:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 174 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:286:

```

Leu His Pro Ala Gly Ala Thr Asn Gly Ser Gly Gln Leu Ala Leu Pro
 1             5             10             15
Val Glu Ala Pro Pro Arg Ser Val Pro Ser His Gly Glu Pro Leu Gly
 20             25             30
Ser Ala Ala Pro Glu Gly Leu Glu Gly Glu Phe Asp Asp Arg Ile Asp
 35             40             45
Glu Arg Phe Pro Val Phe Ser Ser Ala Ser Leu Ala Glu Ala Leu Pro
 50             55             60
Gly Pro Leu Thr Pro Met Thr Leu Asp Val Gln Leu Ser Gly Leu Arg
 65             70             75             80
Ala Ala Gly Arg Ala Met Gly Arg Val Leu Ala Leu Gly Gly Val Val
 85             90             95
Ala Asp Glu Thr Gln Arg Arg Ala Ile Ala Val Lys Gly His Arg Pro
100             105             110
Thr Ile Gly Val Ser Ala Asn Ile Val Ala Ala Ala Gln Leu Pro Gly
115             120             125
Pro Asp Ala Gln Ala Val Thr Arg Arg Ala Leu Gly Glu Gln Pro Gln
130             135             140
Val Thr Gln Leu Leu Pro Phe Gly Arg Pro Gln Leu Ala Gly Gly Pro
145             150             155             160
Leu Gly Ser Val Ala Lys Val Val Val Thr Ala Arg Ser Leu
                165             170

```

(2) INFORMATION FOR SEQ ID NO:287:

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:287:

```

Val Gly Val Val Gly Val Gly Ala Thr Ser Pro Ala Gly Ala Gly Ala
 1           5           10           15
Gly Ala Gly Ser Ala Gly Thr Gly Ala Gly Ala Gly Gly Gly Ala Thr
 20           25           30
Lys Gly Arg Ile Asp Ser Ala Ser Ala Leu Ala Ala Pro Leu Ser Thr
 35           40           45
Gly Leu Leu Ala Val Pro Ser His Thr Thr Asn Gln Arg
 50           55           60

```

(2) INFORMATION FOR SEQ ID NO:288:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 133 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:288:

```

Met Ala Asn Thr Gly Ser Leu Val Leu Leu Arg His Gly Glu Ser Asp
 1           5           10           15
Trp Asn Ala Leu Asn Leu Phe Thr Gly Trp Val Asp Val Gly Leu Thr
 20           25           30
Asp Lys Gly Gln Ala Glu Ala Val Arg Ser Gly Glu Leu Ile Ala Glu
 35           40           45
His Asp Leu Leu Pro Asp Val Leu Tyr Thr Ser Leu Leu Arg Arg Ala
 50           55           60
Ile Thr Thr Ala His Leu Ala Leu Asp Ser Ala Asp Arg Leu Trp Ile
 65           70           75           80
Pro Val Arg Arg Ser Trp Arg Leu Asn His Arg His Tyr Gly Ala Leu
 85           90           95
Gln Gly Leu Asp Lys Ala Glu Thr Lys Ala Arg Tyr Gly Gln Glu Gln
100           105           110
Glu Met Ala Trp Arg Arg Ser Tyr Asn Thr Pro Pro Arg Pro Ile Glu
115           120           125
Arg Gly Ser Gln Phe
130

```

(2) INFORMATION FOR SEQ ID NO:289:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 133 amino acids

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:289:

```

Pro Gly Ser Phe Ala Arg Thr Lys Pro Pro Gly Arg Thr Ala Asp Ala
 1             5             10             15
Pro Ile Arg Cys Arg Asp Ser Arg Gly Thr Ala Gly His Arg Ala Leu
 20             25             30
Asp Glu Pro Pro Pro Arg Gly Ser Glu Pro Ala Arg Arg Arg Ser Arg
 35             40             45
Gly Val Arg Thr Val Val His Asp Ser Leu Ala Ala Arg Arg Val
 50             55             60

```

(2) INFORMATION FOR SEQ ID NO:290:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 72 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:290:

```

Gly His Gly Gly Gln Ser Ala Ile Gly Leu Gly Gly Gly Ala Gly Gly
 1             5             10             15
Asp Gly Gly Gln Gly Gly Ala Gly Arg Gly Leu Trp Gly Thr Gly Gly
 20             25             30
Ala Gly Gly His Gly Gly Ala Arg Arg Trp Tyr Arg Gly Pro Thr Ala
 35             40             45
Ala Arg Ser Gly Arg His Gly Arg Arg Gly Trp Arg Arg Trp Ala Asp
 50             55             60
Arg Gln Arg Arg Gly Arg Arg Arg
 65             70

```

(2) INFORMATION FOR SEQ ID NO:291:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 74 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:291:

```

Arg His Arg Arg Arg Ser Leu Ala Ser Leu Arg Ser Ala Ser Ser Pro
 1             5             10             15
Ala Arg Ile Thr Gln Val Arg Pro Tyr Thr Pro Leu Leu Thr Arg Arg
 20             25             30

```

Pro Ile Pro Val Arg Ala Ala His His Glu
65 70

(2) INFORMATION FOR SEQ ID NO:292:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 174 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:292:

CCGCACGTAA CACCSTGAAT TGAAGGGGAGC CGCTGGTCAT GGGGCGATTC TATCCGTGGG	60
CGAACGGTTA TTGACGGCCC GGAGGCCACT CCGCTGCCAC CAAGTGGTGA CTCAGCGCCT	120
TTTCACGGCA ACCAACGGCG GACACACCAC TTGACATTCG ACAGCAGGGC CGCG	174

(2) INFORMATION FOR SEQ ID NO:293:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 404 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:293:

TCCCAAAACGG GGTGAGCTTC CTTCCCGTGG CGCTAGAGAG TTTTCTCCAC TTTCGGGTGA	60
CGCTCGCGCG SCAGGCGAGC ACCGCTGAGT TCACGCTGCT AGTGGAGGTC CTCGACGGTC	120
CGCTCGCGCAC GATGCGGTCG AAGGCGCTCG SCAGGCGGGT GGTGGCTGTC TTACAGCGCT	180
TGGTCAGCGG GTGGGATGCG GGGTGGCGCG ACCTCGACAT TGTGCTGAC GCGGAGCAGC	240
ATCGGACCGC ACCCGGGGTC TGGGATGTGA CGAGCTCGGC ACCGCGGTC CATACCGGGT	300
TGGCGGAAAT CGCTCGGGCA GAGGCTGACT CGGTGGCGGT GAGTTGGGCG GATGCTCAGC	360
TGAGCTACCG GGAGCTGGAT GATTTGGCGC ACCGGCTGGT TACT	404

(2) INFORMATION FOR SEQ ID NO:294:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 104 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:294:

Leu Val Glu Val Leu Asp Gly Ala Leu Gly Thr Met Ala Pro Glu Ser
 35 40 45
 Leu Gly Arg Arg Val Leu Ala Val Leu Gln Arg Leu Val Ser Arg Trp
 50 55 60
 Asp Arg Pro Leu Arg Asp Val Asp Ile Leu Leu Asp Gly Glu His Asp
 65 70 75 80
 Pro Thr Ala Pro Gly Leu Pro Asp Val Thr Thr Ser Ala Pro Ala Val
 85 90 95
 His Thr Arg Phe Ala Glu Ile Ala Ala Gln Pro Asp Ser Val Ala
 100 105 110
 Val Ser Trp Ala Asp Gly Gln Leu Thr Tyr Arg Glu Leu Asp Ala Leu
 115 120 125
 Ala Asp Arg Leu Ala Thr
 130

(12) INFORMATION FOR SEQ ID NO:235:

(A) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 526 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:295:

GTTTCGACGG	CTACGAGTAC	CTGTTCTGGG	TGGGTTGTGC	GGGCGCCTAC	GACGACAAGG	60
CCAAGAAGAC	CACCAAGGCG	GTGCGCGAGC	TGTTCCGCGT	CCCGGGGGTG	AAATACTTGG	120
TGCTGGGCGC	TGGGGAAAGC	TGCACCGGCG	ACTCGGCGCG	CCGCTCCGGC	AACGAGTTCC	180
TCTTCACGCA	GCTGGCACAA	CAGCGCGTCG	AGACCGTGGG	CGGTTTGTTC	GAGGGTGTGG	240
AGACGCTCGA	CGGCAAGATC	CTTGTACCTT	GCTCGGACTG	CTTCAACACC	ATCGGCAAGG	300
AAATATCGGC	GCTGGCGCGC	AACATAGGCG	TGCTGCACCA	CACCCAGCTG	CTCAATCGGT	360
TCTTGGCGCA	CAAGAGGCTG	CTGCTGTCTG	CTCGGTTTTG	TCAGGACATC	ACCTAGCAGC	420
ACCTGTGCTA	GCTGGGTGCG	TACAACAAGG	TCTACGAGGC	AGCAGGGGAG	CTGATCGGTC	480
CGCTGGGGGC	CACCTGAGCC	TAGATGCGCG	GCCATGCCGA	CGGACG		526

(12) INFORMATION FOR SEQ ID NO:296:

(A) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 467 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:296:

TCTGCTGGCG TCAATCTGCG TGGGAAATTC CTGATCTGAT TACGATCTGAT TACGATCTGAT
 TACGATCTGAT TACGATCTGAT TACGATCTGAT TACGATCTGAT TACGATCTGAT TACGATCTGAT

CCGCGATAGG ACAGGGGCGAG CAAGCCACCG GCAACGCGGG CCACATCGCT GGACACCTCG 420
 AGACCGTACT GCACCAACCT GAAGAGCTGA AACTCGCGG AACGTGCAAC AGCTGCGAAC 480
 AATTGGG 487

(2) INFORMATION FOR SEQ ID NO:297:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 528 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:297:

ACCAAGCGCG AGAATATGAG CCGGGGCAAC CCGGCATGTA CGAGCTTGAG TTCCCGGCGC 60
 CTCAGCTGTC GTGCTCGAC GCGCGTGCT CCGTGTGGOT GCACGCTTTC GAAGGTTTCT 120
 CCGACGCGCG CCATGCGATC CCGCTGGCGG CCGCCACCT CAAGGCGGCG CTGGACACAG 180
 AGCTGGTCGC GTGCTTGGCG ATCGATGAAC TACTGGACTA CCGCTCGCGG CGGCCATTAA 240
 TGAATTTCAG GACCGATCAT TTCACCACT CCGATGATCC TGAGCTAAGC CTGTATCGCG 300
 TCGCGGACAG CATCGGCACC CCATTCTGCG TGCTGGCGGG TTTGGAGCGG GACCTGAAGT 360
 GGGAGCGCTT CATCACCGCC GTCCGATTGC TGGCCGAGCG CCTGGGTGTA CGGCAGAACC 420
 ATCGGCGCTG GCACCGTCCG GATGCGCGTT CCGCACACAC GACCGATCAC GATGACCGCT 480
 TATTCACACA ACCGCGAGCT ATCTCGATT TTCAACCTT CGATCTCC 528

(2) INFORMATION FOR SEQ ID NO:298:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 610 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:298:

TTAAGCGGCT TAAAGAGCGG TTGCGCGCTT TCGCTGCTGT TCGGCTGAGT TCGGCTGAGT 60
 TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT 120
 TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT 180
 TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT 240
 TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT 300
 TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT 360
 TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT 420
 TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT 480
 TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT 540
 TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT TCGGCTGAGT 600

244

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:299:

```

Phe Asp Gly Tyr Glu Tyr Leu Phe Trp Val Gly Cys Ala Gly Ala Tyr
 1           5           10           15
Asp Asp Lys Ala Lys Lys Thr Thr Lys Ala Val Ala Glu Leu Phe Ala
 20           25           30
Val Ala Gly Val Lys Tyr Leu Val Leu Gly Ala Gly Glu Thr Cys Asn
 35           40           45
Gly Asp Ser Ala Arg Arg Ser Gly Asn Glu Phe Leu Phe Gln Gln Leu
 50           55           60
Ala Gln Gln Ala Val Glu Thr Leu Asp Gly Leu Phe Glu Gly Val Glu
 65           70           75           80
Thr Val Asp Arg Lys Ile Val Val Thr Cys Pro His Cys Phe Asn Thr
 85           90           95
Ile Gly Lys Glu Tyr Arg Gln Leu Gly Ala Asn Tyr Thr Val Leu His
100           105           110
His Thr Gln Leu Leu Asn Arg Leu Val Arg Asp Lys Arg Leu Val Pro
115           120           125
Val Thr Pro Val Ser Gln Asp Ile Thr Tyr His Asp Pro Cys Tyr Leu
130           135           140
Gly Arg His Asn Lys Val Tyr Glu Ala Pro Arg Glu Leu Ile Gly Ala
145           150           155           160
Ala Gly Ala Thr

```

(ii) INFORMATION FOR SEQ ID NO:300:

1. SEQUENCE CHARACTERISTICS
 A. LENGTH: 161 amino acids
 B. TYPE: amino acid
 C. STRANDEDNESS: single
 D. TOPOLOGY: linear

11. MOLECULE TYPE: protein

xi. SEQUENCE DESCRIPTION: SEQ ID NO:300:

```

Arg Val Arg Asp Leu Ala Gly Glu Leu Arg His Cys Ile Gln Thr Pro
 1           5           10           15
Thr Ile Ile Asp Gln Ala Asp Ala His Asp His Arg Thr Gly His Gln
 20           25           30
His Arg Gly His Ala Gly Gly Ile Asp Glu Pro Pro Gly Glu Cys Arg
 35           40           45
Glu Leu Gly Gly Lys Leu Arg His Ala Arg Asn Ala Gln Glu His Arg
 50           55           60
Gln Thr Thr His Val Val Val Val Val Val Val Val Val Val Val Val
 65           70           75           80

```

245

```

      100      105      110
Leu Thr Gln Arg Asp Thr Gly Ala Ala Ile Gly Gln Gly Gln Gln Ala
      115      120      125
Thr Gly Asn Ala Gly His Ile Ala Gly His Leu Glu Thr Val Leu His
      130      135      140
Gln Pro Glu Glu Leu Asn Thr Arg Arg Thr Cys Asn Ser Cys Glu Gln
      145      150      155      160
Leu

```

(2) INFORMATION FOR SEQ ID NO:301:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 175 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:301:

```

Glu Ala Arg Glu Tyr Glu Pro Gly Gln Pro Gly Met Tyr Glu Leu Glu
 1          5          10          15
Phe Pro Ala Pro Gln Leu Ser Ser Ser Asp Gly Arg Gly Pro Val Leu
      20          25          30
Val His Ala Leu Glu Gly Phe Ser Ser Asp Ala Gly His Ala Ile Arg Leu
      35          40          45
Ala Ala Ala His Leu Lys Ala Ala Leu Asp Thr Glu Leu Val Ala Ser
      50          55          60
Phe Ala Ile Asp Glu Leu Leu Asp Tyr Arg Ser Arg Arg Pro Leu Met
      65          70          75          80
Thr Phe Lys Thr Asp His Phe Thr His Ser Asp Asp Pro Glu Leu Ser
      85          90          95
Leu Tyr Ala Leu Arg Asp Ser Ile Gly Thr Pro Phe Leu Leu Leu Ala
      100          105          110
Gly Leu Glu Pro Asp Leu Lys Tyr Glu Arg Phe Ile Thr Ala Val Arg
      115          120          125
Leu Leu Ala Glu Arg Leu His Val Arg His Asn His Arg Pro Gly His
      130          135          140
Arg Pro Asp Glu Arg Ser Ala His Thr Thr Arg His Asn Arg Arg Ser
      145          150          155          160
Phe His His Pro Gly Ala Ile Ser Arg Phe Glu Ser Phe Asp Leu
      165          170          175

```

(2) INFORMATION FOR SEQ ID NO:302:

(i) SEQUENCE CHARACTERISTICS:

- A LENGTH: 175 amino acids
 B TYPE: amino acid

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:302:

Lys Pro Val Lys Glu Pro Val Pro Ala Leu Pro Pro Val Pro Pro Thr
 1 5 10 15
 Pro Ala Leu Pro Pro Leu Pro Pro Leu Pro Pro Val Pro Gly Phe Pro
 20 25 30
 Thr Val Pro Pro Gly Ser Met Ala Pro Leu Phe Arg Pro Phe Ser
 35 40 45
 Pro Ala Pro Pro Ser Pro Ala Leu Pro Pro Ser Pro Pro Leu Pro Pro
 50 55 60
 Leu Val Gly Val Ala Ala Trp Leu Thr Tyr Cys Ser Thr Gly Pro Ala
 65 70 75 80
 Leu Asp Pro Leu Ala Val Ser Ile Ala Ala Ser Met Asp Pro Pro Thr
 85 90 95
 Thr Thr Cys Glu Ala Ser Pro Ala Ala Ala Ala Gln Leu Cys Arg
 100 105 110
 Gly Ser Cys Asp Leu Ala Pro Ala Asp Glu Met Met Gly Thr Thr Gly
 115 120 125
 Ala Cys Gly Arg Leu Gly Glu Ala Ser Ala Gly Ser Arg Ser Arg His
 130 135 140
 Thr Arg Arg Cys Ala Ala Ala Ser Glu Ile Cys Arg Leu Arg Cys Thr
 145 150 155 160
 Arg Ser Ser Ser Gly Val Pro Arg Asp Trp Val Ser Pro Leu Ala Pro
 165 170 175
 Pro Leu

(2) INFORMATION FOR SEQ ID NO:303:

(1) SEQUENCE CHARACTERISTICS:

- A: LENGTH: 921 base pairs
 B: TYPE: nucleic acid
 C: STRANDEDNESS: single
 D: TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:303:

AATTGGGCAAT GAGCAGGCAAT AACAGGAGGCT GATTGAACTG AGGCGAGCTC AATACCGGTA 1
 TTGGGAACAT CGGCAGCTTC AACACGAGGA GTTGAATCG AGGCGATTCC AACACCGGGG 12
 AATTGAACCT AGGCAGCTAT CACAGGAGGA TTGGGAACAT AGGCGATTTC TACACCGGGG 24
 GATTGAGCTT CGGCAGCTAC AGCAGGAGGT ATTGTGGAGT GGAATTATAT AGGGGTCATT 36
 TGTGGAGCTT GGGATTTCGA ATGCTCTGKA GAATTGAACT TCTTGNACAA GCTTGGCGGC 48
 GCACTGAGG CCGGGTGAAT GATTGAGTTT AACGCTNAN GAATAACTAG CATAACCCCT 60
 TGGGGGCTCT AAACGGGTCT TGAAGGCTTT TTGGTGAATA GAGANGAATA TATCCGGATA 72
 AATGGGCTAN TACGAAAAGG CGCAGCGATC GCTTCCGAA CAGTTGAGCA CCKGAATGGC 84
 AATGGAGCNC GTTCTTACCG GGCATTAACT CCGGGGTGTH GUKSTTACCC CCACGTNACC 96
 TATACCTGCT TANNSSGCTN BSGGCTGCTT TGGTCTGCTT ATTGCTTCTT GCMCTTCGCC 108
 GCTTGGCTT AGCTATAAAT CCGGNNCC TTTTGGTTT GAATTATTTC TTACNGSCCC 120
 GATCGGCTT GATTATTTTC TGTATTTT

(2) INFORMATION FOR SEQ ID NO:304:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1082 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:304:

```

AATTGGGCAC GAGATANGGG CGCACCAGGG TCCGCAGCCG GCGGGACCGT CGCCAGCACC      60
ACCGGGGTCA ACAGCACCAC GGTGGCGTCC ANGCGAGCGG CCGCGGTGAT GGCGGGCTGAG      120
ACGGCRAACA CCTGCCGTAG CAGTCGGTGC GACTCCCCCG TCGCTCGANC CATGGCCGCG      180
CGGGGTGGCT CGAACANGCC TTGGTCGTCC ACAGCTTAGC CAGCAGCCAA ACCGCAACCA      240
CAACCCACCA CGCCCGCGCC CCGGGANACC TCCGCCATCG KCTGCTGGGG CGANATCCCC      300
CGATCGCTNA CANGATGACC GGTGGCGGAA CGCCGCGCGT GCGTGGGGG AGCGGCTGG      360
GCGGGGCAAC CGCGAACCCA NGAACACGGC AAGCAGTATC ANCGCAACAG CAATTGTCAA      420
GGGCTAAACG CTTACATCC AGGGATCTCG CGCGCCGACA CCGTCGGMTC TGCAGGCGCA      480
CCCCNTCCTN GGGCGGNCAC TCNTCAAAGA TGCGATCNA CAGKCTAGGT CTTGGGCGCA      540
TATGSAAGGN CCCAACGGNT TTAAAGCGGC SAAAAAATC TCCANTGGA TAAATCAGC      600
CGGGGANCCC CCCGTGSCMM NGTCYCGGKC ATTNTTCAAC MGGTTTNACG GCGGKTGCG      660
GCGAACTKGC CAAAMTTAAG KTNNGGGNTY CGGGCGGTA ACCGGCINTK NGCCCCCTAA      720
AAAACCGGNC YTTCTKGAT TAMMACCGN CGCCCAWTGG CGGKTCKTCC CANGNTYAC      780
AMCCYCCGSE MNGGGTGGG SAACCTTCC TSNNGGGTTC NTKGTTSCYT AWMCCCCCG      840
AAACCGSKY GGTGGGCTN WASSAMNCG CMNGYTTCTT TAAAGGCCAN KNRAAWGKYT      900
CGTGGGAAW CTTCAATYC SAAATKCTC CTTCMGSSCN CTTKWRTYN NRNGGGAACS      960
AMWTRYCCIC GWTTCAWTC GGTCCGASM AAACKCTTYT TTTTCGSSC STCCMGSSNC      1020
GGTGNANAN AAATATTTC TCGNRNANKH TCTCGGCTT TTKMGRNRR GMGAACCGG      1080
TS
  
```

(2) INFORMATION FOR SEQ ID NO:305:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 991 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:305:

```

AATTGGCAGG AGTATATGGT GTAAAGCTTT TAGGCGCGGT GGTGAGGTG TTTTGCGAAC      60
CAATATCGGT CGANGTGTG TCTGTAGGCG GTTCCANAA CGGTGGGGG GTGCGGCGG      120
ATCTGATCGG CGCGGCGGTA GTTACCTTC CGGGGCTGT GAGTGGGAT TCGGAATG      180
TCTGTGTGCT GGTGTACCA GTGAAGAAAT GTTGTGAGT GAGCGGGG GTTGTGATG      240
TACTCGAACC GTTGTGGGAA GTTGTGAGT TAATGAAAT TGTGTAGCT TGTGTGAGT      300
TAATGTGAGT GTTGTGAGT
  
```

AWTCCCAACN	CCCKCCAANA	RCYKGGGGCC	CCNCCCAACC	CGGGKGAAGA	WTAATTTAAA	660
CCCYAACMAW	ACTWMMNACC	CNNGGGSCCY	AAMCGTYYNR	AGGTTTTTCT	NAAAGAAASA	720
ANTCGGAAMC	CGGNTSTACC	AAAAASCCCK	CCNWTCCCTC	CRASATTGSC	NCCSAAWKSA	780
AKGCCCCCN	TCSGCNWNNC	CSGCGGKKRT	EKGTTNCCCT	WMRCWMWYTS	GGCCNASCEN	840
CKYYSMYCC	CCCCTCCCCM	CTCCGNKTCC	CCAMCCYANC	MGGCCCCCYM	GKKCCCWKNT	900
YKCCCCCCCC	AMMNNNGGGG	WGACCCCTGG	CCCCMKRRGM	TCCCNANTGA	MCCTCWGNRA	960
MKCYCCNRAR	ANMCCSCNCC	NGCNCRCNN				990

(2) INFORMATION FOR SEQ ID NO:306:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 223 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:306:

AATTCGGGTG	GCAACGCGGG	CCTGTTGSGC	AACGGCGGGG	CCGTTGTTGC	CGGTGGGGCT	60
GTTGGTGGCG	CCGGCGGGCG	GGGCGGTAAC	GCGGGCTGGT	TTGGTCTATG	GGGCGCTGGC	120
GGGTTGGGTG	GTGTANGTGC	GGCGGGGGCG	AACGGTGCTA	CGCCCGGTCA	GGATGGGGCG	180
GCTGGTGTTC	CCGGTGGGA	CRACRGTCT	GGCGCTGCTG	CCG		223

(2) INFORMATION FOR SEQ ID NO:307:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 418 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:307:

AATTCGGGAC	GANGCGGCAA	CGGTGGGAGT	GGGGGAGCT	GGTTGGGAC	CGGGGGGGGG	60
GGGAACGGCG	GTGGGGGGG	GGGGGGGGG	GGGGGGGGG	GGATGGGCAA	CGGGGGGCAAC	120
GGGGGGAGTG	GGGAATGGG	GGATGGGGGG	GGGGGGAGCT	GGGTGGGGGG	GATGCTGGGG	180
GGGTGGGGGG	GGTTGGGAGT	GGGAACGGG	GGGGGGAGGA	GGGAACGGCT	GGGAACGGGG	240
GGGAACGGGG	GGGAACGGGG	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	300
GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	360
GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	GGGGGGGGG	AAATGGGG	418

(2) INFORMATION FOR SEQ ID NO:308:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 143 base pairs
- (B) TYPE: nucleic acid

(2) INFORMATION FOR SEQ ID NO:309:

(A) LENGTH: 1036 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

X1 SEQUENCE DESCRIPTION. SEQ ID NO:309:

[illegible]

(A) LENGTH: 1036 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:310:

AATTCGGCAC	GAGATCATGA	ATAGCGGGCT	GGTCAGCACC	GAAGTGGTCG	GCGATCTCGC	60
GAGCAAGTCT	CGTCTGCTCG	CCCAGCAGGA	GGTCGGGCATC	GATGCGGACA	CCTGCGGATGT	120
CTTGGATGGT	GTTCA GTTGC	AGGTAAGGCC	GACGCGCGAG	CTTTGCTAGC	AGGGTGTCTT	180
GGGTCTTCGC	ACGTGAGGTA	ACCAATAACT	CCGACGCAGA	CCAACCTCGG	CCCTCGATCC	240
GGGTACCAGG	CTCGGCCGGA	GCCAGCCGTT	GTGCCCCCTG	GGCCGAAGGT	CAGCTGCTGT	300
GCGATCGAAG	TAAGAAACCG	CCCATGCCC	GTCCGCCAAGT	ACGACTGACC	GAGCAAAACGA	360
ACGATCGTCC	TCCTTCCGCT	GGGGGTAAATC	GANCCGAGCA	AGTGCACGAC	CCACCAATCA	420
TTGGGATTGG	GCCACTGACC	GACCAACCCG	CTGTGCGACA	CCCCAGCGGA	ATTGGTGTGT	480
TTCCGCGGGG	CGCNAACCG	AATCANCSSG	ACGCGCTCCG	CGAASCANCC	GCATANCCTT	540
ACATANCAAC	CGNNTCTCCG	CCACATTTG	GGGSTMTCG	CCCTCNGCAA	CSSNAAYNCC	600
CCCAATTCYG	AACNAAAAA	TTGGYCCATY	ARNGTYCTCM	CCAAAAACCN	AWTCCCGCTA	660
TCGCCCCGGG	GTCGCCCTYY	NMNAJAJCCG	CCCWAAANCC	CCSGGSCSCC	CGGGTTRWTN	720
CCCTTTGTCC	GCCCNCCSGG	TTGGTTCMCM	GSCMMTNWN	GGGNTGCSCC	CCNCNAAAA	780
AAAAAYCKIG	NCAAATYAAA	CCCKYCMAAA	ASKTGCGSSC	CCCMARCCCG	GGKAAKKWWA	840
ANTTAANCCN	KAAAAAAAWW	NCANMCCCC	NGGCNCCATA	GGKYTTTAGG	GTTSTTNANG	900
ARAAAATMTC	CANATMNSSK	TTTNAAAAAA	ASCCSWAKCC	CCCNHKKKNV	CCAAWKAARP	960
SRCCTTCGGG	TTWNSGGGGG	KHKKMTNCMS	KMNMTTWGR	CCCNCCGCCN	NNTWKCCCTN	1020
TCCNYGGNCC	RNCAGN					1036

12 INFORMATION FOR SEQ ID NO:311:

2. SEQUENCE CHARACTERISTICS:

A. LENGTH: 1060 base pairs
B. TYPE: nucleic acid
C. STRANDEDNESS: single
D. TOPOLOGY: linear

11) MOLECULE TYPE: Genomic DNA

14-00000

AAATTTTGGAA	TAGTCTATTG	TAATGAAAGG	AGCGTACGCT	TTTCTGATGA	TATGCGTGGG
TTTATTTTGA	AGCGCTATCT	TGCGAAAGCG	AAATTAATTC	TTCTCTTACG	TGTAAGCTGG
AAAATGGTGA	TGCGGCGATG	GATCGAACCG	AGCGCGCA	ATTTGAAATA	TGCTGAAATG
AAATGAGAA	TGCGTCAAGG	AGCTGACAG	TGCGGCGAGG	ATGCAAGGTT	TTAGAGCTGT
TTTCTTTTGA	AGAAAGCGGT	TGCGCACCGA	AGATCACCGT	TGCGAAGGCG	ATCGAGTCAA
TGTTTGTGTT	TTATCGCGGT	AACTTGTGCG	TGCGGAAGAG	ATGAATGTG	TGCTTACGAT
TTTCTGTGTT	ATCAAGCTTG	TGCGCTCGCG	GACCAATAGT	ATGAGAGGAT	TTATGCGGAG
TGCGCTTGA	ATCGGCAAGG	TGTTGAAATA	TTCTTTGTG	AGAAAGGAT	TGCTTACGCG
TTAAATTTGA	ATGCTTTTGA	TAAATGTTCA	TGTTTTCGGT	TAAATATG	TGCTTTTAA
AAAGCAATA	TTTAAATGTT	TAATGATGAA	TAAATGTTG	TGTAAGTCT	TGCGGATGAA

AAGGKKKTKC SCMACCCCAA AAANMMNNCN AWHCCCGMGK SARGGGGRNY TTMEAGGGMG 960
 GNYCCCCCW YCGGGGGGNA NAAYAAAAGK NGSNGRGAAT NTNTNTTTTGK RSSSRNKTIT 1020
 TYNTCTTCYN CCNMGNRWWG SRAMNTGKTS NSSGGGSGGC 1060

(2) INFORMATION FOR SEQ ID NO:312:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1040 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:312:

AATTGGGCAC GAGCTTCACC AAAGAGCTGA CATGGCGGGT GATGCGACAT CCCATCGAGG 50
 GCRATACGGG CATGGATGAN CCGAANGGAN TGTGGCGTTC GGTCAACTGG ATTACGCTTC 120
 CCAAGGTGAA ACGCTTTGGG CGGAAAGATG CGACGCTTAA GTTGGCGTTC CACCGTCCAA 180
 TGTGTGTATG GATGCTGGAA CCGCGCTGAC NGATAANGAA TTGCTGGTTC GCCGGGCGACN 240
 ATGGATGCTC CKSTTTTCNC TCCGCGGTTA AATTGCTGT GGCATCATCTG GCAGGCTATG 300
 TTCCCGCTAC RCTGCAGGCC ATCATGGATG TGCGGCTAAC GAANAAGTTA TGACATGGCG 360
 CAAGCGAMTC GGGCATSCNC GCGGCAMTTT CGCAACCTGC TGTGTNTGAA GCCTMTCAAC 420
 CGAATCCGGC GCTYAAAAGT NGGCTTGGCT TGATTMMAAC CNAACCCNTN CNATYCTTTC 480
 CCGNCGMNTG CGTTCTCTCC AACTCGGKKG GTTGGCGCCG TGAAACCCMA CTNCCCCCCC 540
 GTTGGACTTA MRTNTTCAAA AAMCGGNTNA ACCCGAATN SAACCTNCCR TCAAAANTAMM 600
 SAANTCGGGC TTYGGGNRCC CCGGGAAYW TTGNCNGGG GMNNTYCTCN GGTTYNGGG 660
 SAAACNTTTC CORTNCGMN TTTACAMGGC NCMNTMTTGM GGGSCSNNAS GWCCCCGGKK 720
 TMTTTCNAW TONENKTTT TTKGGGGGGG GGTGGRTRMC NCGGGCCCCC GGCCCKKMAA 780
 AAAACCMCA RRCNCGGG KKCCCGCCCM NMATNGGGCC YKCRAAACAA ACCCCANRA 840
 TNGNGMGGGC SMACCGNGN BYHAAAGGT TSNSTMANM MKGMANNCT SGMSCCMNSN 900
 NCTGMSGGKT TTKGNAGRN AANAMKMGGM RSGNCGGCRN GAAAGGGGMS GSCKSCNNGN 960
 NGASNGWMN TRINGANRCC NONGNBMN NUNNNNNNNN GGRNNNAON NMKMCAMSMC 1020
 NSRNGNNNL TGYMTNKGCC 1040

(2) INFORMATION FOR SEQ ID NO:313:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 144 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:313:

AATTGGTAA GAAAGAAAT GGTCAACTGG GATGCGACAT AACTGAGGCT GGTGGTTCAGT 50
 GAGCTTCACC TCAACTTTC TGTGGCGTTC GGTGTATG GATGCTATG ATGCTGGCGG 120
 ACGCTTAAAT GATGCTATG TGTGGCGTTC GGTGTATG GATGCTATG ATGCTGGCGG

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 332 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:314:

AATTCGGGCAC	GAGAGACCGG	GTGTTGACC	AACGGACGCT	TGGGCGCGGG	CCCTTGGCGT	60
GGCATCAGCC	CTTCTCCTTC	TTAGCGCCGT	AACGGCTGCG	TGCCTGTTTG	CGGTTCTTTGA	120
CACCTTGCGT	ATCCAGCGAA	CGCGCGATGA	TCTTGTAGCG	CACACCAGGC	AGGTCTCTTCA	180
CCCGGCCCGC	GCGCACCAGC	ACCATCGAGT	GCTCCTGCAG	GTTGTGGCCC	TGGCCCGGGAA	240
TGTACGCCGT	GACCTCGAAC	TGACTCGTCA	CTTCACGGGG	GCAACCTTCC	GAAGCGCTCG	300
CTTCGGCTTC	TTGGGAGTGG	TGGCTCGTGC	CG			332

(2) INFORMATION FOR SEQ ID NO:315:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 962 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:315:

AATTCGGGCAC	RAGTCGTCT	AGACGGATTC	AATGCTCCCG	CGAGCACCTC	GGCACTGCAC	60
ACCTTCGAGC	AAAATGTGCT	CAATGTGGTG	AACGAGCCCT	TCCAGACGCT	CACCGGCCGC	120
CGCTGATCG	GCAACCGCC	CAACGGGACT	CCTGGAACCG	GGCTTGACCC	GGGGCCGGCG	180
CTGCGCTGT	CGGCACCGCC	GCGAACGGCG	GCTCGGGGG	GAACGGAACC	AACGGCGGGG	240
AGCTGGGAC	CGCGCCGGGG	GGATTTCTTC	GCAACGGGGC	ACCGCGGGGG	CCGGCGGTGT	300
CGACAAAGCG	CACCGCGGGG	GACCGNGCCG	CGCTNGGGCG	GCTTCTKGAT	GGGCTCCCGC	360
GGTACCGGG	CACCGCGGGG	CGCGGCTGAC	CGCTNGTTGG	GACCGCGGGG	CGGCTNACCC	420
CGATCTTCTT	CGCGNCCCGG	GAAACTCGCG	GGCGCGCCCG	ACATTAKACC	CGCGCGNACC	480
CGGMCCTGG	CGGAACGGNG	GGTNTTTTCC	AACGGGGGGG	CGCGGAACC	CGMGGSTGTT	540
CGTTNGGCGA	AGGNCBAAT	CGGKCTANG	CGATTCCTCG	ANGGKTGAMC	CTGATSSNCA	600
CGTTMAGGAA	CTTNGCCANT	CTTNGGACAT	CTTNGGAA	AGRAWNKNGT	CGGCAAACTA	660
CTTNGCTTCT	NATKGSNNA	AAAATCTCTT	CTTNGGACAT	CTTNGGCGNG	CGGCTNCTTN	720
CTTNGCTTCT	CGCGGNAAM	CTTNGGACAT	CTTNGGACAT	CTTNGGACAT	CTTNGGACAT	780
CTTNGGACAT	ANGGKSNCG	CTTNGGACAT	CTTNGGACAT	CTTNGGACAT	CTTNGGACAT	840
CTTNGGACAT	NGKKNKNA	CTTNGGACAT	CTTNGGACAT	CTTNGGACAT	CTTNGGACAT	900
CTTNGGACAT	NGGKSNCKT	CTTNGGACAT	CTTNGGACAT	CTTNGGACAT	CTTNGGACAT	960
CTTNGGACAT	CTTNGGACAT	CTTNGGACAT	CTTNGGACAT	CTTNGGACAT	CTTNGGACAT	962

(2) INFORMATION FOR SEQ ID NO:316:

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:316:

AATTCGGCAC	RAGAAGACCC	CCGAANGTTT	GCGCTGGCTC	TACAACTTCA	TCAARGCGCA	60
GGGGGAACGC	AACTTCGGCA	AGATCTACGT	TCGCTTCCCC	GAAGCGGTCT	CGATGCGCCA	120
GTACCTCGGC	GCACCGCACG	GCGAGCTGAC	CCAGGATCCG	GCCGCGAAAC	GGCTTGCGTT	180
GCAGAAGATG	TGGTTCGAGG	TGGCCTGGAG	GATTTTGCAN	GCGACGCCNG	TGACCGCGAC	240
GGGTTTGTG	TCCGCACTGC	TGCTCACCAC	CCGCGGCACC	GCSTTGACCT	CGACCAGCTG	300
CACCACTCGT	GCCGCTCGTG	CCG				323

(2) INFORMATION FOR SEQ ID NO:317:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1034 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:317:

AATTCGCAGT	GTGTGTGGCG	GCSTCCAGAA	GAAGATGATC	GCGAACATCG	CCAGCGCCCGG	60
CCAGGCTATG	GTGCCGGTGA	TGGCCGACCA	GCAGATCATC	AGCGGCATAC	AGCCGGCCCGC	120
CCGACCCAC	ACCACSTTCT	GTACCTGCG	TGCTTTGAGC	CAAAGCGTGT	AGACRAACAC	180
ATAAAACCG	ACGGTACCA	GGCCAGCAC	CCCGCCAGC	AGGTTCGTGG	CGCACCATAG	240
CCAGAAGAAC	GAGATCACCG	TGNACGTAC	CCAGTGCCTA	ACCGSTTTCC	GGTCGGGCACC	300
GCTTCCCGCC	CCAAGGCGCG	CCCGCGCGTT	CGTTTCATCA	CGTTGTGGAT	ATCGGCGCTCC	360
GCACCACTTT	GAGCGTCTTG	CCCGCGCGCG	CCGCGCATCAT	CCCGCGCGACN	ANCGTGTGTA	420
GCATGANCAG	CGGATCAATG	CCCGCGCGCG	TGCTGCGCGT	CGTCCCGAAT	TCAACTCCCT	480
GNACAACTTG	CGGCGCGCGT	CGAAGCCCGG	TGAATGAATG	AATTTAAACC	GSTSAACANT	540
AACATACATA	CGCTTGGCGG	CTTTTAAACG	CTTTTGAANG	GGTTTCTTCC	TTAAAGGAAG	600
AACHATTTTC	GCATATCTCG	CTTTTWTATC	GAAGAGGCCC	CGCGCATNGC	CGTCCACAGT	660
TTSCCGCTGA	ATGCGAATCG	GNCGCGCTTC	CGGGGCTTTT	AACCGCGGCG	CGNTTTTGTG	720
MGCTCTCTCA	CTTCTTCTTC	ATTTTCTTTC	CGGGGCTTTC	TTTCTCTTTC	NTCCCGCTST	780
TTCTCTCTTC	CTTCTCTTTC	ATTTTCTTTC	CGGGGCTTTC	TTTCTCTTTC	NTCCCGCTST	840
MAAANASAAH	CTTCTCTTTC	ATTTTCTTTC	CGGGGCTTTC	TTTCTCTTTC	NTCCCGCTST	900
CTTCTCTTTC	ATTTTCTTTC	CGGGGCTTTC	TTTCTCTTTC	NTCCCGCTST	960	
CTTCTCTTTC	ATTTTCTTTC	CGGGGCTTTC	TTTCTCTTTC	NTCCCGCTST	1020	
CTTCTCTTTC	ATTTTCTTTC	CGGGGCTTTC	TTTCTCTTTC	NTCCCGCTST	1080	

(2) INFORMATION FOR SEQ ID NO:318:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 141 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

```

AATTCGGCAC GAGCCACAT CCGGGGCGCG TCSTTGCATG ACTCGTTCGT CATCGTCGAC      60
RAGGCACAGT CGCTGGAGCG CAATGTGTTG CTGACCGTGC TGTCCTCGGT GGGGACCGGT      120
TCCCGGGTGG TGTGACCCA CGACATCGCC CAGCGCGACA ACCTGCGGGT CGGCGGCCAC      180
GACGGGTCCG CGCGGTGATC GAGAAGCTCA AAGGTCATCC GTTGTTCGCC CACATCACCT      240
TGCTGCGCAG TGAGCGCTCG CCGATCGCCG CGCTGGTCAC GAGATGCTCG ANGAGATCAC      300
CGGGCCGCGC TGAGTGGGCC TCCCGCGAGC A

```

(2) INFORMATION FOR SEQ ID NO:319:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1026 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:319:

```

AATTCGGCAC GAGATGCTCA CCGTGGCGAC CAGTGCACCG AGGCCACGCC ACCAGTTACG      60
GCTGATGGGC CAGAAGATGG ACCAGGTGCT GCCCATCCCG CCCACCGCAC TGCAGCTGAG      120
CAGCGGGATC GCGGTGCTCA GCTACGGCGA TRA3CTGGTG TTCGGCATCA CCGCTGACTA      180
TGACGCCCGG TCCGAAATGC AGCAGCTGCT CAA3GGTATC GAACTGGGTG TGGCGCGTCT      240
GGTGGCGCTC ANCGACAATT CCGTCTGCTG GTTTACAAGG ATCGGCSTAA GCGTTTCATC      300
CGCGCACTCC CCANCGCTGC GCGGCGGGGG CGGCCCTCTG TGCGGACCGC CGGAGCGCGT      360
CACTGACGCG ATCTCGGTG GCGTTAACCC CGTGAGAAGG TGGGTGCTGC GCAAGTTGGG      420
CCCGGTCAAC ATCNATCCCG GCCGCGCATGA CCGNGT3CTG TTCCACACCA CNTSNGACNC      480
CCCCCAGGAA CTGGTCCGGC AMTNCAGGAA NTYCGT3TGG GCACENGCTT CTTCCGKTRT      540
GGCYTAAACT TCCNATSTTN CGGCGGGCCT CTGGCGTTNC GNCGGGGCGG NTCTTNCCAA      600
ATCGGMMMAA ATCCCCANMC AAACCGCCCG GGTCTT3GG GC3GGGNGGC GGCNAWNCC      660
AAACCGCCCG NTAAANTCT TTGKTCNNN CNGCGGNC NCAANSCAN CCGTTTGGG      720
NCTTCCCGCC CCANWTTTAA CGGAKCGSCN AAYCGCAAY TMMGKCCYCY KNAAAAAAAA      780
AATTTGCGCG CCGCAANTAA ATTCCGNGG CCGTTGGGGG CGRANCNTNT TTTMCCSNSS      840
TGNRRAAME NGGANC3GG KAAVTMMTNS AAYCGCCSN AAMNTTTTC TAANNCCCN      900
TNC3CGGAAA ATTNVAMAAM CTRKTTGNG GGGKTTGNC SGKKGRAGGM AAAAAANRSN      960
SKTTTMCNNN SANMNCNSNN GGGNSNNNNN NNNCNCYKC CNAANMCCC CGCGGGGGGG      1020
CGMMCC

```

(2) INFORMATION FOR SEQ ID NO:320:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 324 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:320:

GCACCACTCG TCGCGCTCGT GCGG

324

(2) INFORMATION FOR SEQ ID NO:321:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1010 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:321:

```

AATTCGGCAC  GANCGGTGCC  GCTNAACACC  AGCCCGCGGC  TGCCAGATAT  CCCGGACTCG      60
GTAGTGGCGC  CCGTGGCGTC  GTTGCTCTCC  TGACGGGGCG  CGGCGACCAT  AAGGTGCGCTM    120
ATGCCCCAGGT  AGCGGCGGAG  GTGCATGGAG  TCGATGATGA  TGGGACTCTC  CAGCTCGCGC     180
ACCGGGAGCT  TGGCATCGGG  CCGGATCAGC  CAGGACCGCT  AGGACAAGTC  GATCGAATGC     240
ATAGTGGCGT  CCAGAGTGGC  CGTGCAATTC  CAGCGTGCTC  CACCGCAAAT  GCCTTGATTT     300
CTACTCGCGC  TANTGTTCCG  GCATCGCGTC  CGGATGAAT  GGGAACCGCA  SGATGGCGAC     360
GAACGGGTCT  GANCTCAGGT  TTGCGGCTTT  GCGCACAGTG  GTCNACANCC  GGTACTCGGC     420
ATANATCTGG  CCCNAAATCG  GCGCGGACGG  CGGCCACNAT  AANAACGGGC  ACNACAATCG     480
CCGCCCCCGT  CACCCNAACA  ACANCTTGGC  ATCGGATTTT  GTCCCCANCG  CTCANCCGT      540
CCCGAACGCC  TONTCCGGCG  NACTTTTCTT  NNAWTAAGTG  CCGCTTCCGK  CCCTGGNGCA     600
WTAATGGGA  AACCTTNCG  CCACCTTGAA  GGGGTTGTTG  NATTTTACT  GSTAACCCCG     660
AATTTTCCG  GANTCGGTCT  KCGGGGTTTT  YSTNTTCCCC  ACCTTNGNAN  GGGCCGGCCA     720
AGSTTTTCTT  SYTGAAGGGG  GAAACCCAA  TTNTTYTYYN  AACGSCMNA  MYMTTTYCSG     780
MNAASCTNKT  CCCCCTTAAC  CAMGGSGGTN  AACCGKTMNG  NGGKTAAAA  GGGSKNNKTG     840
NCGCCYMANG  GGGGCGAAAA  TSTKTCNCG  GGGCCKAAAW  ACCMMMYGN  GTGKKKNKSS     900
TCSAAATTTT  NMMRAACTN  GGGGCGSSGA  NNTTTNAAAG  MSCGCGCSN  GSTGKCCCN      960
NTTTCORAA  NMKKGKQWNM  SNMNSCGNG  SKGSSGSNN  NNAAGMGGGG      1010

```

(2) INFORMATION FOR SEQ ID NO:322:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1010 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:322:

```

AATTCGGCAC  GANCGGTGCC  GCTNAACACC  AGCCCGCGGC  TGCCAGATAT  CCCGGACTCG      60
GTAGTGGCGC  CCGTGGCGTC  GTTGCTCTCC  TGACGGGGCG  CGGCGACCAT  AAGGTGCGCTM    120
ATGCCCCAGGT  AGCGGCGGAG  GTGCATGGAG  TCGATGATGA  TGGGACTCTC  CAGCTCGCGC     180
ACCGGGAGCT  TGGCATCGGG  CCGGATCAGC  CAGGACCGCT  AGGACAAGTC  GATCGAATGC     240
ATAGTGGCGT  CCAGAGTGGC  CGTGCAATTC  CAGCGTGCTC  CACCGCAAAT  GCCTTGATTT     300
CTACTCGCGC  TANTGTTCCG  GCATCGCGTC  CGGATGAAT  GGGAACCGCA  SGATGGCGAC     360
GAACGGGTCT  GANCTCAGGT  TTGCGGCTTT  GCGCACAGTG  GTCNACANCC  GGTACTCGGC     420
ATANATCTGG  CCCNAAATCG  GCGCGGACGG  CGGCCACNAT  AANAACGGGC  ACNACAATCG     480
CCGCCCCCGT  CACCCNAACA  ACANCTTGGC  ATCGGATTTT  GTCCCCANCG  CTCANCCGT      540
CCCGAACGCC  TONTCCGGCG  NACTTTTCTT  NNAWTAAGTG  CCGCTTCCGK  CCCTGGNGCA     600
WTAATGGGA  AACCTTNCG  CCACCTTGAA  GGGGTTGTTG  NATTTTACT  GSTAACCCCG     660
AATTTTCCG  GANTCGGTCT  KCGGGGTTTT  YSTNTTCCCC  ACCTTNGNAN  GGGCCGGCCA     720
AGSTTTTCTT  SYTGAAGGGG  GAAACCCAA  TTNTTYTYYN  AACGSCMNA  MYMTTTYCSG     780
MNAASCTNKT  CCCCCTTAAC  CAMGGSGGTN  AACCGKTMNG  NGGKTAAAA  GGGSKNNKTG     840
NCGCCYMANG  GGGGCGAAAA  TSTKTCNCG  GGGCCKAAAW  ACCMMMYGN  GTGKKKNKSS     900
TCSAAATTTT  NMMRAACTN  GGGGCGSSGA  NNTTTNAAAG  MSCGCGCSN  GSTGKCCCN      960
NTTTCORAA  NMKKGKQWNM  SNMNSCGNG  SKGSSGSNN  NNAAGMGGGG      1010

```

AATTNTTCCG	GANTCGGTGN	KCCGGGSTTT	YSTNTTCCCC	ACCTTINGNAN	GGGCGGGCCA	720
AGSTTTTCTT	SYTGAAGGGG	GAAACCCCAAC	TTTNTYTYYN	AACCCSCMNAA	MYMTTTYCSG	780
MNAASCCKNT	CCCCTTTAAC	CAMGGSGGTN	AACCGKTMNG	NGGKTAAAAA	GGGSKNNKTG	840
NCCCCYMANG	GGGGGRAAAA	TSTKTCNNCG	GGGCCKAAAW	ACMMMMYGN	GTGKKKNKSS	900
GCSAAATTTT	NMMRAACTKN	GGGGCCSSGA	MNTTTNAAAG	MSCCCCCSNN	GSTGKCCCN	960
NTTTCNNAA	WMKKGKNWNM	SNMNSCSNGG	GKYNSSGSNN	NNAAGMGGGG		1010

(2) INFORMATION FOR SEQ ID NO:323:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1092 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

XI. SEQUENCE DESCRIPTION: SEQ ID NO:323:

NSNGGGGWN	NTCAYCAYCA	YCACSGGGW	GNATTGCGGG	CGCAWCTTGT	MAASAGATCT	60
CSAAYTCGGC	AMGAGGGAMT	CKCTMGCNCC	GCTGTGCAAN	CCAATRAGGT	CTRATAATTY	120
CUACTCCACA	AAAAACCGTT	GTGTGTAYYT	SCCGRAAATR	AAGGCGCCGG	TNTCAACWYC	180
GCUGGTATTY	CCRATYCCCG	TKTTGTAMCT	GCCKGGGTSR	AAAYCCCCGG	TGTTGGAYCC	240
CGGATTGAA	ACTSCCGGKT	TGAAACTGCC	GKTTTSGCSA	TCCGGKWATT	GAMSTCRCCG	300
ATTAAAAAAC	CGGKKTGGN	GCTGSNCSTG	CCAAATNCGR	AYCCRATAYC	CCATGGCCCTG	360
KYCTYCTCK	YCGGTACCCA	AAVCTGGGTA	TGCTATACTG	GYCCCTAAAK	GCAAWYCKGG	420
GCTSYCMETK	TTGEXGGSOT	CCNAATTTAS	CACCASGGGT	TCCTTCGATA	CCNAAACNCG	480
GKTGGGWCW	AGMCCGAAA	AAAKAATAAT	RAKAAKGGTG	CATNYCCAAA	ACCNCGCCCN	540
CCCNANTNCH	ATCCGNTNCC	MSCNCCCCCA	GGGGTNAAGK	TKSGGAAYTT	CTMMAACCCC	600
CAAANCCCCA	TAACNTNCGR	GAASAAAACC	CTYCNCGGGG	GYCNWNCAAA	ACASNTTTAT	660
TTGCTKSTTT	CGGGMWCCGT	GCCCCCNAAA	YCCCAASTA	CTTTTGGGGT	CCNAGAKAAA	720
ACNCGGGGN	CCMCCSNAA	NWTATYTCTT	KGGCAANCCG	CSAAACCTTR	TCMNACCNCH	780
ATPMTCCGTT	CCCTTSCAAT	TGGYCGGRAT	NCSSNCCYTY	TCAAARKKKS	CAKWNWNGNG	840
BRPNACOMA	ACCGAAGTY	CCNAAAATN	JKCCCCGCTG	CNAACAGNKH	TYTTCGAAA	900
AGCCCCCCCC	CCNCCCCRAA	AACCCCCCNA	CKANTNCCCA	AAAACNPNKH	GGCCCCCCCC	960
CAAAACMAAAA	AMCCCCCGSM	AMACCGGGGN	NMCCCCGKKK	KKTTTCTCTT	TKCCMRSCCG	1020
AAMSCAMWSY	KSKTMAAAAA	GGAAURANCN	TYCCANANM	TCCCNVRSW	CCGSWGMGNA	1080
GAASMCCCCC	CS					1140

(2) INFORMATION FOR SEQ ID NO:324:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1251 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

XI. SEQUENCE DESCRIPTION: SEQ ID NO:324:

ATATTCCAGG	CACCACGCCC	AGTTTGGTGG	ACAATGCCCT	GGCAKTTTCC	TCRAANTTCG	300
TGAAACCGAA	TTTNSMTTGA	ACCNCCAARG	CCCCSNCCNR	AACARTTGGG	WTCCGCGGTT	360
CTCCCCACCG	KTTTCCGGGG	GTNTCGGCAN	AANCGCACCC	WTGGWTTCTM	TCNCCGCACC	420
GGGCGGACAA	NTCGGTTGC	AATTTGCRA	AYCGGGGCGG	GGATTCCSCA	AACGGGTGCC	480
GAAACTGTTY	YCRAAMACCG	GGAKCCGCAA	TTTCCGGGCR	ANAAATTTCTN	YCNCACCACT	540
GCTTCTACTT	CCCCGACCGT	AACMANTTTC	ATCGTCTNNN	CCTCTGCCCT	TGGGGCAGGG	600
CKAAAYACCG	CMTTKGGTTT	CGCAACCTGC	CGCCCAANTC	CCNAMCCRCA	CTTTCNATTT	660
GGNTCGAATT	SCCCCCCGGT	RANAACCSCC	NTGGCCNNYT	CGGASSAAAA	NGGGCCCTNT	720
KGGCNSCCCC	AGTAANACCC	TACCNAYTS	CAWTCTTTGC	CAAASITKGG	ACGAANSKTG	780
GGNTTCCGGK	ATTTYTTTGS	GGNCNCCCTN	TATNGGSNTN	GGGCCCKCYN	NCSTKTGKCA	840
NASSKAYCCS	NGNKGGGGGT	ACCCCCCTMG	GGGGGTTTTT	NSGGCCCCCC	AWAYGNKSTG	900
GCCCCCNNGG	GGAAKAATWT	MWWTMCNSGG	GGGAAWTTTT	NTSTGGAMCS	SGGACYCCCR	960
GGGGGKTTTT	TCCCCCNCSA	NNAWANGGGG	GGGGGANAYT	NTGNSGNGGG	KWNTTTATTT	1020
YTYCYCCTM	TKACMSGGGG	GTTTTKAKNG	GGGGGAGAAA	ANAAAAAANA	RAKGGYKNTT	1080
TSKNCACNCT	GKWNWNWANR	NAGAGTCCCT	CKCKCCNCSC	SNTTTCTTTT	MGNSGSYGGG	1140
GNNGNNNAAA	ACNKSRRMAC	KCSYTYCCCG	CGYCTCCTCC	NCNGGGGYGS	NGSGGNSTYN	1200
BNMKGRKMTA	TNTMGNCGTN	SCUTCCNCCC	GCKNKNTGTC	TMTCNMYGSG	C	1251

(2) INFORMATION FOR SEQ ID NO:325:

(1) SEQUENCE CHARACTERISTICS.

(A) LENGTH: 1099 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(12) SEQUENCE DESCRIPTION: SEQ ID NO:325:

AAATCGGCAC	MGAGTATCAC	CAAKCTGYGT	TGCCCCAGCAA	AGTGGAGCTA	TACTACCTG	60
TATGTGATCC	TCRACATCTT	CTCCCCGTAC	NTGGTCGGGT	GGATGGTGGC	CTCGCKTGAK	120
TCRAAGGTCT	TGSCCRAACG	GCTGATCGCG	CAAAACCTTG	CGCCGAGCAC	ATCAKCGCCG	180
AACAGCTGAC	CTGCMCCCGG	ACCGGGGGYC	GNCAATAACT	CCAAAACCGT	GGCMCTGCTG	240
TTGGCCNACY	CGGTCTCCCA	ANTCGAAGTC	ASCCGGGNMA	CCAKMAACKA	NAACCGTTGT	300
TTGAAGCCCA	GTTCAAAAAC	CTRAAGTWCC	GGCCGACTT	CCCGAAACGG	TNCGAGTCKA	360
TORSAGGSGG	CGGGGTGCMC	TGCAACCGGT	TCTTCGGNTG	GTRCACCCTN	AAAMCAAGCA	420
TTCCGGGMITC	CGMMTCCCCA	CGCCGCCAAS	TTTMCCTACG	GCSGSCCNAT	CAAATTCGCC	480
TGGAAACGSGN	CGHCCCTCNK	CGAMACCGCC	TWCCAAAAGC	TCGGAACGGK	ATCCTTCKGY	540
NAACNCCCGA	RNCCTCKSKT	TCTGGGCTTC	NMSGGGAATA	TCCCKNSCMNT	CGGAATCCAA	600
TTCCCMYFGG	TTTTTWTYCC	CCCGGGCCCG	AAATNGGGYC	CTASSNMKC	ENCAMNANT	660
GNWATCTGG	NGTTCCUNAN	TTTGGGCTTC	NMAATGAMNA	NMNRGGGTTC	TSCYACCMN	720
NACCGHMHG	KCCCKKCTK	MANAAKATC	RATCAMKWNG	GGNKCKCNCN	NAAMACCCCN	780
ENCYNWYTC	TYCCKSKW	CGMYNANCA	NGGGGAGGW	GGSGRMKMC	CTMTCTCNC	840
MCCGGCCTNT	TYCKSGAKAT	ACASMNKTC	CGCCNGCCCN	MAAMANRAKA	CTAKCCGYGN	900
CCSNSTCTTN	CTSNMKMNR	TCCWMWNATC	NTTYGKKCNC	KCTMKATNWC	TSCTSKCNCK	960
MRAMTCTYGG	SNMTCTCTCA	TCTCTCKKCC	SNMSKCTCKC	KSCNCCNCWN	CNKCNMKCWN	1020
GGNSTCRCTY	TCTMNNCTCS	AGCKCCSKNC	WACNCACACK	NGWCTTTTCC	WKNNNMKCNKM	1080
TCKCKCACPR	MTMTWCCG					1099

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:326:

GNGNTATACA	TCWCTGTGYA	CCSAGGATCW	ANTGCGGGCG	MAAKCTWSTM	CASAGATCTC	60
AAAYTCTGCA	MGAGCGGCAC	AKAKYSTCGT	CCMRACCCGG	CAYACWCCWG	CNCGCCCCWT	120
CTTRGACCGG	GGCKATASMC	ACCGTTGGCC	CCGGCNCGCA	CCTACACCAC	CCACGCCGCC	180
AGCGCCCCCW	TRAMCAAACC	ACCCCGCKTT	TACCGCCCGC	GCCGCCGGGG	CCACCACCAG	240
CCCCACCGGC	ACCACCGGCG	CCGCCGTTGC	CAAAACAGGC	CCGCKTTTGC	CACCRA	296

(2) INFORMATION FOR SEQ ID NO:327:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1073 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:327:

NONGSSNKMV	ATCATCWTTC	TGCACCSNGG	MTCWATTGCG	GCCGCAATCT	TSTMNASAGA	60
TCTCGAAYTC	GGCAGGARCA	TCTGCGCGGN	GAATGTCCAA	AWGTCKWTKA	CGGCMATCGG	120
TTTGCCGYCA	ACCACKCTRT	SCAKATGCGG	GCCAMWTYCA	AACCRATTAT	TTGGGYCGAG	180
AAAATTTMCG	CKTGTRASCA	ACCTGTAGCG	GSTCAASCAA	CAGCCTCTRA	ACCGTAAATY	240
CKTAGGTNKT	YCCGGCAACA	ASCVCRATAA	TSCGGCCCGC	AMCCACAAAA	CCTGANTNGT	300
TNTTCNCRAA	NCCGGTYCCC	GRAGGGGTSA	ACTGCGGTAR	GCTTNTCWYC	NCCTTRACAT	360
TAAACCCGCG	CGGNTGWTCS	CCGCGGCCAA	ATYCTTSCCC	WTKGCNACCA	YCCGANCCTG	420
CGGTATGGTS	RAANCASTGG	GCRAACGGTM	MRCSTACCKC	TGGGTGATYC	KTGGGNTCCG	480
SNAAATCGGG	GATTTACGGG	LAMGGTTAAV	CGAGYCCGGG	TNTGCTCKY	CNACACCCGG	540
ATCMWNCNCG	TACCTKTTAA	AATTCCTTGT	GGTGGAAAGC	AWYCKAAAAA	NMTNTYCCGN	600
TCCAMMGGGG	CYCGGAAKYT	CNACNTGGKT	NACCCCTNCG	YTTGAASTTT	TCYTGNCCTC	660
GGGCCAAAS	ANACCSGAKC	CCCGGAAYCS	WTAGGCVTCN	TGCCCCSTTA	AATTGKNCYT	720
AATCCCKCAA	CGCTCCCGCG	GGTCCSCCMT	TAAAMTTCCG	CCCKSCASNG	GAATYCYKGG	780
GNSTMATTV	CCNCCGNTTT	TYTGHNAAAJ	SCCCGCKGN	TGCTYCCGN	GNTTSSGCGG	840
STTTGAMYC	AAAWNTNGG	MMCNRAGNCG	SGNAMCCGTH	TKKSGGGAAT	TKAAYTYGGR	900
GGGGTCNYC	CCGRCGNAA	AAGYCTHGG	KCCGSSCCYT	TCMARTTTYT	CNGGMRCMAV	960
ATMANSNGNS	CTCCGSTNCW	NGCCTTCCGN	SNMAMAAAAA	WYKCCGNGG	TCGARNRGNV	1020
MTT LKNGNG	WTCCCKNKTG	NSCNGGNCOS	YCGNSASWCC	YNYCNGCACA	ANC	1080

(2) INFORMATION FOR SEQ ID NO:328:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1166 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

CGCCCCGTTG	TTMMMTTCAY	TCATTCACCG	GGMTCTAGTG	CGGCCGCAAK	CTTGTCACAA	60
GATCTCGAAY	TCGGCAGGAS	ACAATSTCGG	GTKGGGCAAT	GTCNGGTGGG	GCAACTTTTG	120
GCTCGGRAAT	YCGGGGTAA	CGCCGGGTCT	RATGGGTSTG	GGTAATATCG	GGTTTGGTAA	180
TGCCCGCAGC	TACAATTTCG	GTTTGGCAAA	ATATGGGTGT	GGGCAATATN	GGGTTCGCTA	240
ACACCGSCAS	TGGRAATTTC	GGTATTSGGT	NACCGGTRAY	AAYCTGACCG	GGTNCGGTGG	300
TTYCAATACC	GGTAACGGGA	ATGTSGGTTS	YYYACYCCGS	GSAACGGNWW	YTTNGKTCCT	360
TMCNCTSSM	CKSAAMTSM	KMGSTSTYCT	MTYCNNGGAS	TAMTYNMCCC	CCGWAYCKSC	420
WAYCCCTCGT	CATYCCMCMC	SGSGYCCTCA	MNCCACCCTG	NGYYCCCTCC	MKMTTCYCAIT	480
CMNTCCGGTW	CCTNTMMNCC	CSCNCRYCTC	AMCNCTKSGK	CACCNATMYC	CSACKCHTCT	540
MCYMCSCAKN	MTTCCCTCN	CCTYTNCCA	MCMCSCTCTM	TCMAACTCKC	CCGGYCKCNC	600
MYCTCTCKCC	AYNMAACCKK	TYCYWNCWYC	YMYCKCKCAG	WYKNTCTCCW	ACTCTMYNTT	660
TCTCTCNKCC	CMKACCKNTT	CTCWCSCCCC	CCACAKAYMC	YAWCMTMTCC	MCTCKACSCC	720
CYYCNMYCCM	NMCWCMTCWC	TWNAKCANCN	TTCTTCTCTC	MMYMTMACKC	WCNNTCNCCK	780
SGACCYTCTC	ACTKMKCKCK	TCTCCTTMCK	CCYMWCTNCC	MKYNCCCTCC	NMTCTCKYKT	840
CCTCNMRY	CYYAKCAK	NMCTCCCAN	KMCAKCTKCT	CCCCCAKMK	ACNCKCCCHC	900
CCTCTATCC	WCTCTCWC	ATCTCKCTCW	CNYCMYMKMC	ACNCKCYAYT	CNACTMMNWN	960
CCANCTCT	CTNYCTCCK	ACCTYCKCK	CTMCKCHYMC	NRWCTYRCCT	CKKCCNCCRN	1020
CKMCMKCTM	CTCTCCWMM	TCCWCSCAT	CTMMKSTCTC	WCNCTCCCT	CHKCCYNYNT	1080
KCYTYCCMYG	CTTCKNTCT	MCCWCCYATC	TCTMKCTCT	CWCACYMCAC	WMTTACWNCC	1140
ACTCTCTRCW	CKCKCKMCCR	MTCTCB				1166

(2) INFORMATION FOR SEQ ID NO:329:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1230 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

11. MOLECULE TYPE: Genomic DNA

xi) SEQUENCE DESCRIPTION: SEQ ID NO:329:

NGNGGNNNT	CWTACATCWN	TCTNCACCS	NGMTOWATTS	CGCGCCGCAW	NCTTGTMNAS	60
AGAATCTCAN	AAVTGGGCAC	ANATGTCTTT	TSTMTAKTGT	CGCGGGGNGC	CACGCCXTAT	120
GTGYGCTGG	GYTPACGAA	CGCGCGGCS	CGGGCCRACC	AGGCGGGGRA	TSCAGGCCGC	180
CGCGCGCGCG	CGGYTATAT	RAAGCGCGGY	TTTTKTRATA	ACGGTSCCGC	CGCGGGGTRA	240
TTACGGGCAA	AAVCGGKKT	TTGGGTATAT	AACGCTAATT	GCAACCAWTT	TTTTYCGGGT	300
AAAAACGCG	CGWGANATC	CGCGGYCMCT	GAAGCGCAT	TCGCGCAAAA	WTNTGGGCGC	360
AAAAACCGCT	TCYTATTTT	TGGGTATCT	CGTGTCTTG	GCAACGCTY	CGCGGGTTAA	420
TGGCTCCCG	CGCGCGCGN	AAAAACGCG	AATYCGGYT	CGGGTGKTC	CMCAGGCGGT	480
TGCTYCGNGY	CGCTCGGCA	AAVYCGCAW	AKATTGGGTG	CTYCKTSCG	TTCTYGGGCT	540
CAATTACCG	TCGCGNAAA	AAAAAANAA	ATGNTCTNT	TGCTCGGYCA	YCTTTMTTGG	600
AAAAAGGGG	ATGCTSCGGT	TYTTTACCT	CAAYCGGCA	NCANTWACCT	YTCGSCCGCG	660
CGCGNANAA	CGSTTNGCT	CGCGNACCT	TCGTCGCGN	ATCNAAAGGC	CNGAATTTGG	720
TTTSTTYCS	ATTWTWKHY	CGCGWNTT	AAAAAAKCCA	AAASAKCCCK	YCNCAMMYKT	780
NGGGGTYSS	CGCTYCTTH	SNMTTAAAC	CYCGGCAAAA	YNSGGGKKT	TCCGCTNSAT	840
KCGGCTNCG	CGGGGCGNA	AAAAAANAA	TTTTGCGAAA	ATCCGACCTY	TCYTKKSTY	900
AMACCGCTT	TYMKHAYT	CTYCNATCT	CGCTTGNAAA	TYCGGYGGCT	TNTTCCCGCK	960
CGGGGCGCG	AAWTCTCTT	CTYCNATCT	CGCTTGNAAA	TYCGGYGGCT	TNTTCCCGCK	

(2) INFORMATION FOR SEQ ID NO:330:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1022 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:330:

```

NGNGGGGKNA  TMAYCWTCTC  ACSSGGTCTA  TCGGGCGCAW  CTMGTMASA  GATCTCNAAY      60
TCGGGCAMNAN  GCATMTCMMC  CATATATAAC  CATTCGCTCS  GYWTSCAWCT  CRAAWCTGTC     120
CTTCGKGCCG  TTKTACRAAG  GTGGMWTGYT  CWTYCCTRAA  SCCTCTCATC  TCKTKTATYC     180
CTKGGGCTYC  ACTTTAACSG  RATKSCTGCC  TTKTAYCATT  RATGCAAWTA  WTGGYCRAWT     240
KTTSCAGGCC  PACGGCWYCT  TTTTCGGCRA  GRACAATNGA  TTGGAWYCSC  TYCGCRAGGC     300
CCGGCACCAR  ACCGGGCNCC  AAAGGYCCGC  GCAAWTSCCT  GGKTCAAAAA  TGGTGC AAAAC   360
AAAMCNATCC  CCGGYTTRAC  CGCAGYTAMC  ACAAKAAAAA  TCCCWTGGCC  GCACCAWNNT     420
TTYCRATCWY  CWYCCCCACC  TTRAACCTSK  YTGCSGTATT  GCCTKCCTGC  CTCRACAGCM     480
YCNCCCKTCA  AACCTGCGGT  GACTCCAACT  GGTCTGGYCG  AASGGGGGYT  CAMCGGACAA     540
AACCCCRANN  TCGCCAAATT  TTNCCCCGCC  CYCGGGAAN  GKTGATMTTC  TCSNAACCSA     600
CMGGGNNTW  NAACCCGTAA  CSSSGSNKGA  MYNSCCSGGA  ANTTTTCCCT  TYNGGGCGRN     660
AANCCCTTTT  AAGGTACCCC  KGGNGGGGKG  CCCYTTTGGG  AAAACAACCC  CKATTGGKTT     720
TGGAAATNTT  TKNCCCCCA  TTCNSGGGGG  GGGCCCCAMC  CCMCTTTTNN  TCMSCNMITYY     780
YCYTGGGAAT  TNYTCCCCSG  GAAYYCGGSM  CCKGYCCTAA  NCTCCMNWGG  GKYSTGSNAR     840
GGRATMAWWT  TYSTTTVYMC  CCGGCNNCCC  CCCAKMONT  KONTGAACMA  AAKCSCGGGG     900
GSTRMYMWY  YCNNGNNTT  TNRGGSSNMT  TYMAAAMMAN  GGGGKTWYTY  CKCCNGSCNN     960
SKTSGGGST  TTTCCNTTTS  GGGSSATYKS  MACCCCKTMT  AYCGGGGGGT  NTYTKYCCCC    1020
37

```

(2) INFORMATION FOR SEQ ID NO:331:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1088 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:331:

```

NNCGNNHTTA  TAMAYCWYCT  NCACCSGGGA  TFWATTGCGG  CCGCAATCTT  STMAASAGAT      60
CTKAAATCG  GCAMGANGCG  TAWGTATTTG  KGTGRASCCG  ACCAGCGRGA  CCTGCGSGKT     120
GKTTTCTTGC  AGRGAGGCGG  TGGGTGGGCG  CGGTGGCAAT  GCGAACCGCC  CCGCAAAACN     180
CGGCAATTCY  CAAAAACAAA  CCGSGGGGTA  SKTCCSGGCC  TCCAAATMAA  TAACCGTKTT     240
AALNCAGGCG  ACCGCGCAAT  TNYTCTTTC  AACGNAGTNA  TTTCCCTCCC  NATAGGYCCG     300
TTGGGGGCTG  CCKTATYKCG  AANTCTTAY  TGNACGGGM  GGGYCCMONT  TCGGCTCAT      360

```

SACTGCTTTN	TTGNGGGNAM	ATANATYYTT	YCKTNGGGSK	TTCCGMCNC	AMMAATNTCC	720
RGGGKAAMCC	AGKNTNNTCC	YFYCCCCAA	NNTYCCYKGG	RMCTNNYYCY	TTAAANRASR	780
SAACCCCKSGG	GKCYNCNCSS	TARCCCCCAM	KAAAATTTCC	CCCSSKTTTC	TYVNNKQMRW	840
GCCCCCSAAM	ACTMTWAYTT	TCCCKCGNMN	TTTSYCCCKCS	KCAMWMMWMTG	KKNCITTTTTT	900
YCSCMATAMA	CTTNGGKCCT	NTCNYGSGCG	CMAAANAAGG	CGCGSTTCTN	TTCCMMAMACA	960
YNTSGNMMMA	SAAKAKWATA	AWNNTKRYK	TKNNCCCNCC	CKCKCTTSNN	TNKCCMCSKS	1020
GGGKNWNKKR	GWCTCCWCNC	CKCCCNCKNK	CKKWATMCCC	CCCCSKCCGM	NCMMNTTTKT	1080
CCC						1083

(2) INFORMATION FOR SEQ ID NO:332:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1069 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

11. MOLECULE TYPE: genomic DNA

(3) SEQUENCE DESCRIPTION: SEQ ID NO:332:

GGGGNNKYAT	MCAYCWTCT	YACSGGGMNC	TATTGGGSCC	GCAWYTNGT	GASAGATCTC	60
GAAYTCGGCA	MGAAAAAGW	GATGTGCTGG	ACCTTMCSCC	GCGGGACGCR	ACCRACAAAG	120
RAASCSCGCC	ANAATATTGG	CCACAKTTGG	TCACATATTT	ACCCAATTMT	AYCAGGGGAYT	180
MGCAATTCCKG	GGACCRACCG	CACAATCCCR	ATSKTGGTTT	GCRAACCCCT	ACCGTCCCCA	240
MYTYCGCCRA	STTGAACCAG	GGCRAAAAAA	CGGCRAAWY	CTGCCCCCTGA	NTCCCGCTCS	300
GSCNAATAA	CTAGGCCCAT	TKAACGGAAC	CGGNGGCCSC	NANTTGGCCA	ACAGGTCCCT	360
ACAAAGGGGC	CCCASYCGG	CGGGWTCSCW	TTYCACNCCC	TNKTCTCKTG	CCGAATYCGG	420
WTGCRATNYC	CCWTGGGCTT	TKTGWYCKYC	KYGGGTNECA	AWTCTNGGTA	TNCTATRGKG	480
TCCCCATAAT	GCANATCTGG	GCKYCCATTT	NCTGGGNTTC	NATTTAMMAN	SRRCGGTTCT	540
TTGWTTCORA	AACCCSNTGG	GCTTNNMCCA	AAAAATGATN	ATAATAATGK	YGSCTTTCAA	600
AGCCCGCCCC	CCCATTCTRW	CGTTTCGANE	CCGNGNGST	TAAGKTGGGA	ATTTMTNAMC	660
YCNARGCCCT	NATTTGGGNA	AAAAATYCYE	GGGYCTCAAA	EMNYTTTTTT	GSKSSNTCCG	720
CTCCTTTCCC	CAAAACCCAA	ATTNTGNGG	GGYCKTNAA	ACMCGGYCRC	RCCGGAATT	780
TTTTTGSTTC	AACCCCAAGC	TTTTTAAGCC	NTTTTYYT	TGCCSSCSMN	TNGSGGGGNT	840
KGGCCNTTCY	RARKKCCNMN	GGGGGWYCYN	CCGCMNTTT	CTTTTTTTT	CCGTNNMAAM	900
NGKTTCTTCA	AASMCCCCCC	GGCCCTNSAA	ACCCCTNAR	STTTTYCMAA	AANNWYNNGN	960
KYCCCCCCCC	MMNAAAAAAV	YCCCGGNNRN	ACCMNNGGA	MGCCCGGSON	NTTAKTTTTT	1020
CTNMGTYCCG	CGRMASHTTT	TYAMAMANER	HAMHMTTTH	THNPNWNH		1080

(2) INFORMATION FOR SEQ ID NO:333:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1010 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

11. MOLECULE TYPE: genomic DNA

YCCCGGCAAA	GGGGCCACAA	CCTGCAGRGA	SCACYCRATG	GKTGYTGKTS	CNCGGGCGGG	240
CCGGKTNAAG	GGACCTGCCT	GGGKTGCSG	TMCAAAATC	WYCCGGGGGT	YCGCTGGRAT	300
MCNCAGGGGT	GTCAAAAAAC	CGCAAACAGG	CACSCCANCC	NTTACGGGS	CTTAAAANGA	360
AAAAAGGCTG	ATGCCCCCAA	GGGGGCCCCG	NCCCAACCTT	CCGTTGGTCA	ACAACCCGGT	420
CTCTCKTGCC	RAATCCGRWT	CCRATNYCNC	CWTGGCCTTK	TCKYCTYCTY	CGGTACCCAA	480
ATCTGGGTAT	CCTATASTGT	CCCCTAAWTT	CCAAATCTGG	GCTGTCCATT	TSCITGGCWT	540
TCCAAATTTA	CCANCAACGG	TTTCTTNCAT	NCCAAAAACC	GNTKGGCKCC	NRACCCRAAA	600
AAATGAATAA	TAATAANNNG	KCNNTTYCNA	ACCNCCCCCC	CCCNATTCCA	IYSNGTTCCA	660
NMNCCCCCAG	NGGKTAGGTK	GGGAAANYYC	TCMACCYCA	ANCCCTWARS	TTTTNGRAAT	720
KAAACCCCTC	YCNGGGTCWW	TYMAAAAAMA	NTTATTTGGN	NGNTTTTCGG	MWNCKRKNST	780
SCCAAAATCC	MAAATANTTT	YYTGGTYCNA	TWAAAAAMCG	YGNCMMNCCC	GGAAAAWTTT	840
TTNTGKTTSA	ACCCCAAAAC	YTTTTCMNAA	NCSSKTTTTY	CYTTCCCCCC	AMNWTGGGYS	900
GGGNATKGYG	SCYTNTCTTA	TKTKYTYMTW	CMGGGGGGNN	MKMTICMMCC	CCMTTTYCY	960
NYWRTTTTTN	KCCCCKTNMR	MNRAANNNGN	YTCSYNANAA	AAGCNCCCCC	SCCKNCCCN	1020
AAAAWCCCN	NNNARAKTNT	TTMKANNRMN	SKCNKNGKY	YCCCCCCCWC	YNNMNAAAAA	1080
AATMYCCNCC	RASANMCASM	NMGGRGNRSC	CCCCCCCCST	NNNNNTMINT	TTTTTCSRA	1140
SAGCKCCSCG	MNHNAMKNC	CTTTTTKNC	MNGNNGNGNN	GGNGMNCKCC	CCNAGAAMWF	1200
CTKSTCCCKS						1210

(2) INFORMATION FOR SEQ ID NO:334:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1105 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(12) SEQUENCE DESCRIPTION: SEQ ID NO:334:

NGSSSNENNA	TMCATCWYCT	SYAGSGGGMT	WATTGCGGG	EGCAACTNGT	MAASAGATCT	60
ESAAATGSGC	AAKANACACC	ACCGCGCTGT	MTATACACCG	AAAATTTTCT	GTKTGCCAAA	120
AGGAGAGGCG	GGCGGCGCGT	GGSYTCCAA	GGKTTACGTR	ACCGGCGAGY	TCAGTGTTRA	180
AGCGGCTGYT	RAGGSGGSCA	CCCAACWTAA	ACGCTTTAKC	TAAGRAWYTG	GKTGGCGCGC	240
AGCGACCTGY	TGTGTYTGGC	CTCWYCGGTG	GTAGCGCGCG	TTANGCGCGG	TTGCGCGYTC	300
AMCASGSCGC	CGGTRATCCC	AKCNWTCGCC	CGGCGMRACC	GACCGGGCAC	TTTGRACGGT	360
GTGCGCAATT	CAAAAYCKYCT	GRWTCCTTCY	AAACGCCACG	AGGCGGACGM	CTMSCAGCNA	420
ATMGSGRACT	TTAAGGCGCA	GGCAAAACCT	NTKACNCGCT	CGCGGCGRAA	GGTGGCGCAA	480
ATPATCGMAA	AAAANKENIAT	TTGCGCGCAG	AKCAAGCGCA	CGCGGCTTTCT	TTGTTTCCGA	540
TTGGAAMGCA	ATTMCWGGGT	NCNNGGCGAA	AGCAAGCNCG	NWTAKCGTGG	TCGCGGGGCA	600
ATTTGCGPAA	GAAGGCTCTY	CGCGGCTTTT	ATGCGCTCMG	CGCGAANAAT	CGCGGGAATC	660
AAAAAGGGTG	CGNCAAAANG	CGMAAACCGC	GACCGMACTT	NTTCCGCTCT	GGGCGGCGCN	720
CTKMTTTTAA	AWKSGCTGYT	TTGCGCAAA	TCGGCKMAAA	MNGRKTGGGK	TTNGGCGNACC	780
NTTTTGGGGK	CGGGGKGGK	WGKYCTMNMA	CGTTTNTTTT	CGCGGYKAAA	NYSGCGCGCT	840
CGCGGCGCGG	CGCGGCGGGA	MNTTTTATA	GKKTTCGCGT	CGCGMAAAAA	ANAAGCGNYT	900
CGCGGCGCGT	TTKRWAAAMN	KUTSGCGCGG	GNNGGGGCKM	GKTTATTMT	NNNCGSCCGC	960
TCGCGGSAAA	AAATAKMTTT	SYCGGCGCNC	CTCGCKNCKR	HKAMSMCGCG	TCGCTCTGCT	1020
TCNNTTAAAN	APGNTCTYCN	CTNNTCTGCG	NANNTNWCOD	NCCTGCGCT	NKCGCKNCKN	1080
MAAANAATTC	NGSMSTGSMY	CTYCT				1140

- (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:335:

NGSNSNKNMN	TAMAYCWYYC	TSCACSNNGA	ACWANTGCCG	CCRMWCTNS	TMKASAGATC	60
TMGAAYTCGG	CAAGAGCGGC	AAGAGTGTGT	GCATCTGGTC	ANAGTSTMMA	CRCGGTGCCG	120
CSGGTGKGT	GASCACMCAT	NTGCCGRACAC	CAAACCCCKTC	GCGGGYCACC	GGCKTCGCCCT	180
GCAAAYCCT	CCAGGCCACC	TCRAACAAYW	YCTYCTGCAA	CGCARGCCGT	TYCGCGGCCG	240
RATCCTGGKT	CASYCAGCC	TGCCGTGCCC	AAGKTACTGG	CSCAYCAAAA	CCGCTCCGGG	300
RAACRAACKT	AAWYTGCCG	AATTTCTNTT	CCCTGCCGCT	TGATAAATTT	NTNAAGCCAC	360
CGCAAMCCTY	CGGGCKTCTC	CTCKTGCCRA	ATYCGRWTCC	RATAYCGCCA	TGGCCTNKTC	420
KYCTYCKYCS	GTACCCAAAT	CTTGGGTATC	CTATANTKYC	CCWAAANRCA	AWTCTGGGCH	480
KTCCATKTSO	TGGSKTCRA	ATTTAMMACA	NCGGTTTCTT	TCWTACCAAA	AACCSNTGGG	540
CCCCRACCRA	AAAAAGATAA	TAATAAKGTG	CWWWC AAAAC	CCGCCCCCCC	RRTTCAAYCG	600
GTCCARCACC	CCANGNGGTN	AGGTNGGAAT	TYTMAACCCC	CAGCCCATAA	SNTTNSGNAA	660
AAACCCCCCN	GGGYMYCAAA	AMMCTTTTTG	GGGMTTCSGS	CCATKGYKCC	AAAACCAAAA	720
TMTTTCYGGT	CRWAAAAACC	GGCCCNCCCC	NAAATTTTTT	GKCAACCCCA	AACCTTTMAM	780
CCNNNTTCYV	YCCNSACAA	TNGGSGGNKN	NGSSCNTTYT	TWTTTYYNNA	GGGGGGRRWC	840
SNCCCCNAAN	YYCCNAANKG	NKCCCGSNMA	AAAGAGANTT	YCMKAAAAAC	CCCCNCNCCC	900
NAAAYACCCC	MAAAKWTTCM	AAASMSNNNG	YCCCCC			936

(ii) INFORMATION FOR SEQ ID NO:336:

(i) SEQUENCE CHARACTERISTICS:

- A) LENGTH: 1042 base pairs
 B) TYPE: nucleic acid
 C) STRANDEDNESS: single
 D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:336:

NRNGNKNNNY	ATMMAYTCWY	YCTCCACCCG	IGNNWCWATT	CCCGCCRMW	KCTTGTMMAAS	60
AGATGCTNAA	NTCCGCACAG	ASSGCGACAG	ASCCGCGCGG	TTATYCMYCC	NYTGGTCAATG	120
ITCAACACCG	TCYTCGGGHW	GRATATATGG	NGSSGCGCGG	GGCAACACG	YTCAAATGCT	180
TGCCCCACCG	GATATNTTAA	AAAGSTRATA	AAAGCAAAA	GGGCGCGCGY	GGCTTTGGGG	240
JCGGAAACCG	GTCCCAACCG	AAAAAGNCTT	TGGGAYCCTG	KTGRACCTTA	AAGGTAATG	300
PKTCCCTGCT	GGGTATGCT	GGGCAACAAA	CTCTYTGCGG	WGGGTCTGGG	CTTGGGYCAC	360
CCYCRCTTTT	TATMTNCTCY	CTTCAAGNCT	TKGSTYCAAG	CAACCCACTT	CACMAAATTS	420
TTTTGGGKTS	GGGSSGCGCG	YGTGTRNCCG	TAATAATGCG	NTGKTCSGGC	MYCACCGGWA	480
GCATANGCTG	GCGCGGCTTG	GAAGATTTCC	GAATGATATY	GGTTCTGRAC	CCCCACAMRC	540
GTNGAAATCC	GRATCAATTC	CCCNKGGCTT	NTCYCTCTON	GTGCCCAATY	TGGTTTCTAT	600
RKTNGCCYAA	TSCAATGGG	TTTCURTTCG	YGTTCGAAN	TTNACAAMAS	GGTTTCTCMT	660
ACCAAAACCG	NTGSSGCHNA	TCNAAAAANA	RAAANAANKG	KCTTTYAAAC	CCCCCTCTAT	720
TCWYCGGTN	CMRNWCCNNA	NWAAASXBN	GAAGTTTTHA	CCCAANCCMT	ARSTTSGNAP	780
AAAACCCNCG	GGGTGCAAT	GGGTGCAAT	GGGTGCAAT	GGGTGCAAT	GGGTGCAAT	840

(2) INFORMATION FOR SEQ ID NO:337:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1073 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:337:

```

NNSGSGMKKK ATAMATCWCT CTSYACCSNG GMTCWATTGC GGCCEMAWTC TNGTMAASAG      60
ATCTCGAAYT CGGCCAAANAK ACGCMAYGTC AAGTGTRAYY CGGTCCACATA TCMTCCGCGNG      120
TCAACMCCAA AGCCGNGTCA CCGYCTCCCT GGGGCGCCAC CCGCATCGGT RATGCAACYT      180
CGCGCGCCAC CGYCAAAAGG KTCWTTTRAGG CGCTAAAGGT CAMCAATTCC TRAGGTTCMN      240
CACCGTTNTT TGGCGCGCCC RAWTYCTRAC CCGCAATWTC GGTAAATCGGR AATTTGGGCW      300
YCGGCTTGGG CAATAAGKTN TTGGGCAACG GCGGRWTCYC NCTGGCCCGRA ATTCCNCAT      360
TCCKTTAACS SKTGRACCGT TTYCCCGGYT GCGTAAYTG YTYCMTGGGC GCCYTCGGCC      420
CRNASCASY YCRCTAACGGY CMCCAGGCAA TACCKTTGSC TTTTAAACCAC CGGRATNAAY      480
TCKTACCCAC YTCALSSGTS CTGRANTTRK TNTCNTGAA AANMCCACCN AACCCGGNTT      540
RATCTGCTTC MTCANCWTTT SCGGGTTCTT GCGGTTTTGR AAYCTTNATC CMTYCAAAAG      600
GTTTAMTTTC CCAANRAAT CCGYTTGCCA CCGTGGCCGS GGCTGGGTTT CGMWCCTTRP      660
AMATCCNCCS GCGGGSAAAN AMTTGGNTT SSSCCGGTCC CCGGNAATAT YCMTGGNCCT      720
GNAAATTGSS GGGATCCCN GSGNAYCCGG CCWTKGGGGK TNCCTAGTTG GWACAATTYC      780
WKCCGTTCCA AACCGGGGNC CGGGGCGTGG SSCCCTTTT CCTMYNNAAA AAGKGTITGN      840
NYTTTTCCG CNRAANTTCA CCKCNKNTT GGNCCNAACY YYYCAANTTC CANACCTTTA      900
AASAAANCYK YGKTYCCCG TTTTMCSSGS SANCSCCCCM NMSSKNCGGG AAAAAAGNK      960
ELNGCCTTAN CTSNKTNTT TNKTYCCCGC NMWNNSNMCM NCBKHCNKRY NGNSNMNCCT      1020
MKYKCNNNN SRRNNKCSN SSNCGMKYM CMNCCNGMK NGNKSNNCCC MSC      1073

```

(2) INFORMATION FOR SEQ ID NO:338:

i. SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1061 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

ii. MOLECULE TYPE: Genomic DNA

xi. SEQUENCE DESCRIPTION: SEQ ID NO:338:

```

ENNGNHNWEN TMCAYCWYCT SCACSGGRTT TANTCGGGCT GCATYTNCT CKASAGATCT      60
GGATTTGGGC AMNANAARTG TGCTCTCAA TTTAAKKTS GTGKTAAAY GGCACAGGC      120
GNSACCRACA CCTGNGTCA CCGAAAANA CAAAGCWTG AATWTCAGG GCGRAGGCGC      180
TCTCAATYCC CRASCATTA ACCSTTCTG TAAAGGTTC CRAACNAGG ACCCAGYTCA      240
CGCTCGGCA AMTCCCGCTB CGCTCGGCTN TAAAGTBAF TYTGAGGCT RWTCTCTGG      300
TGGYDAMCHT GGTAAAGGCG CWCCHGCHN AATCTTGA TTTCTCTG TAAAGTBAF

```

TCTNTYCGGT	GGGGGSGGGR	ANMYTTCTCT	YCCNNAASAN	OTTAMYCCAN	TTGGSSNTCC	720
CGGKCAAWS	NGGGGGGGNA	AAGGGGCCCC	CGGNTSCKCU	GGGGKKGCCC	CYGGKTTCAA	780
AANTTTCSGG	GKTSTMSCGG	NVTCSCCCCC	CSGCCAAGRA	CCGNGGTTTT	TTTTTGAACC	840
KCMANTCOSA	AMCCGCCSSC	CCMAAAGGS	GCCTNAAWGR	RAYTTNKSCC	CNNAAACSGG	900
CCCCCAKYTY	SGGKTTCCNC	CNCCSGKRG	CCMISTTTMM	MRCCCTTTGN	GNKTTTTTAN	960
MGSCCTTNNC	CACCCCTYCK	GGGKCSMNNA	GAATMYWKC	CNGGGGNAN	RSCCCCCENN	1020
SGKGGGGKG	MGAGYSCCKT	CTKGCNCN	YKNTTTCCCC	C		1061

(2) INFORMATION FOR SEQ ID NO:339:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 986 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

11. MOLECULE TYPE: Genomic DNA

X1. SEQUENCE DESCRIPTION: SEQ ID NO:339:

GNNGNNNKWN	ATMCAYCWYY	CTSCACCSGG	GMTCWATTGC	GSCCGCWKY	TNGTMAASAG	60
ATCTMGAAAT	CGGACANAG	CGGCACAGAG	TGTGTGCATC	TGTETCANAG	CTGTCAACGC	120
GGTGGGGGSG	GTGGTRASCA	CMCATTGCCR	AACACCAAAC	CGTCCCGCG	GYCACCGGCK	180
TGGCCTGCAA	AAYCCTCCAG	GGCAGTCORA	AACAAYWYCT	CCTGCAACSC	ARSCCCTTYC	240
GTGGCTGCRAT	CCTGJKYCAS	TTGGCKTGC	GGTGGCCCAA	GCTACTGGCS	CWYCRANACC	300
TCTYGGGGR	ACENAACGTA	AATCTTGCCN	AATTTGCNTT	CCCTCTSCCC	TTRATNAATT	360
TCTTAAACCA	CGCAACCTY	CGGGCKTCTC	CTCKTGCCRA	WTCCGRWTCC	RATNYCGCCA	420
TGGCCTNKTC	KYCTYCKYCS	GTGCCCAAT	CTTGGTATCC	TATATTGTCC	CTAAATGCAM	480
ATCTKGSTTS	TGCATNTGCT	GGGCTTCAAA	TTWAMANCAG	NGGTTCTCTY	CTTCCNAAAC	540
CGSTTGGGCC	CAAAACNAAA	AATGATNATA	ATAATGGTCC	TNTCAAACCC	CGCNCCTATY	600
CNATCGSKCC	AMCCTTCRGN	GGKTANKKGG	SHAATTCTMM	AACCCCAAGC	CATAASNTTG	660
CGANAAATCY	NCNCGGGYCA	CGAAAACANY	NTNTTGGNY	SSNTTGGGMN	YCATGGSTNN	720
CGAAAACCCA	AATACTNYYS	GGYCAATAA	AAMMSGGYGC	SAMCCGGAAA	WTTTTYTTEN	780
CTNAAACCCA	AAKCTTTTTT	CNAACCCDAN	KNTYCTTNTT	CCPDMANTGG	CNSGGARTYK	840
SSCTTTCACA	ATGKYCCMAA	AGNCGGRAVA	CCARCCGAAA	TTCTTNNNTN	KNHGCTCNST	900
CTNAAAAGGG	GGTTCICMAA	AACGCTCTCT	NCNCTCCCAA	AAKAMCCCCN	AAAGATNTCN	960
NAANASCTCN	NNNSCCCCCC	CCMMMN				986

(2) INFORMATION FOR SEQ ID NO:340:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1074 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

11. MOLECULE TYPE: Genomic DNA

X1. SEQUENCE DESCRIPTION: SEQ ID NO:340:

ATMACCAGCY	CGCCCAGGGY	CACCTTGCCC	AAAACTCCT	GGGTGAGCCA	AATTTCGCS	360
CGGGCCAAAC	ACCANCCGCA	TYCTGGCCTC	AATCYCACCG	GGCCCGGTGY	TAAAMMANMA	420
GRATCTCKTC	MANCCCCCAN	TCAGCSYTNA	CNGCMACAGC	CGCCCTTCTT	CAMACCGCCA	480
RTACCGGGWT	CAACCGGCCS	STCAAACCTCA	ACAGGCGGNC	AGGCCTCCCC	CGGANSAAAG	540
GTCTTACSCC	NNYAANAAAA	MAAGNTCTGT	TTTCCCCCTC	CASAASNAAA	AANCCCCSGC	600
CGGGCCCTTCN	NMMGGGTTTG	GGGMANANAA	AARCNC CGGN	GGAACGNATC	CGAAAMCTCC	660
CAAGTCNCMT	TWAWAACYCYN	NNAACCCCCC	ANTTTTGGGA	AAGGNTCCCC	NTTMYCCCCC	720
TTTTASGKTS	GGGMMYYCTY	TAAAAAAATT	CCCCAAAAAG	CCCCGGGAG	GGTCMAMCTG	780
GGNAAATTTT	CAAMCCNWGK	TTNTTYNGGT	TMCGGGGGRA	AATTTCNCTC	CCYNNNGGG	840
CSSGSNNNAT	TAYGGMSNMT	TTTNNAAWTM	NSGKTSAMM	YNNKCCMNNN	SNNMSMANNK	900
INAMCKCCCN	CCTCNGNGKY	CSCYNCCCCG	GNAGNGGRAS	MKCCNANMAA	AYASGNTTNK	960
CGGAAMMCNN	AATKGNNSC	CCGASMCNM	NNNMAAATMT	CNCNKNSNN	AANRGMRACT	1020
CCCNNSNGMN	RRGAARMTNY	YCCCCGSKM	GKGNKAAAAG	GKYCCCCCM	AAAG	1074

(2) INFORMATION FOR SEQ ID NO:341:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1195 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Genomic DNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:341:

NGNGNCHKNT	MTACATCWT	CTGCACISGG	ONTOWANTSC	GGCCGCWAWY	TTGTCSASAG	60
ATCTCGAAYT	CGCCAMGAGG	ACWCTGSGPA	CGCCCCCACA	NACTCTGGCG	TGTGTACCCC	120
ATTGNGGGGK	TCACCGCGCC	AYTGANCCAF	TNACTGGGG	TGCCSTYCCG	CKTGCGCGGC	180
GGCTCGACGG	CKCTSCWTCT	RAAGGCWTTG	TGCACUGCAT	TGGGTTTTCT	RAACGCTGGG	240
AAAWTGGCCA	GCCGTCTGGC	TCATGGGNTC	TACGCAACSC	CNGCCCCCAA	CRCTTTCTTA	300
AATCCGGYCC	NTCCTGANC	CTTTGAAYCC	GGGGGSAAGA	ACTGGTTGCS	CNCGAYCTSC	360
TGGAACCTTK	TCNAAATCCC	GCANAKTGT	TCNTAMGYCC	CNCCGGAAGG	NGAACCTACT	420
TTGNGGWANG	TGCGCNCCK	GGCCTTATCA	GTCTGATCA	ACGGGGAAC	GGYKNNSTTG	480
KGGGAAAAAG	RRCTCAATC	MTYGGTCCCK	GTCTGKANC	CGCCCCCTGK	GYCCGNAATG	540
GAAGGCSMAG	GGTTAANGCC	MTTYCNWCCP	RSCTSTGTA	GGKWTTYCCG	MGGANKAMNN	600
MNKAMMWTTK	TCRGNSGCCW	ATSTSCCGGG	TKSTTAKAGA	ANACTYCCW	WCCGTNTYCC	660
AAAAGNTRCS	CGCMGTTTTT	CCCKMGANGN	YCTGATTTGA	GGGGGKYKCC	CCCCGGGTTC	720
TGAANKWKYK	CGYAGGGGGM	NTTCCAGTCC	TCMNTATNAG	AGNAAAGKTT	RYGSTSHNCC	780
TYTHKGGACT	NSCCHNWSAK	ANAACNNKNT	TCCCGCNTMS	AGNKTNIKGT	YCCNKTCTTC	840
TAAGAGGAGT	TATYMKCSCC	CTTGGANGMM	GAJWGMGCTT	TYTCCCNKRT	TCNTNSWAAA	900
TATKSAGMGT	TYCCSMAGMK	TCGCTTTTCT	CTTGAANAAMN	MSMRKNNKTS	CGMGYTCTCC	960
GGNTTTGTGA	IAGTARTCCS	CCGCSMWGAC	WCGGCMCMGNG	AGHNTKTRTTS	YANTGARGCY	1020
MNNSKTMKMT	MSGSGSUNA	GGAGNGCCCC	CGANGMSTGY	NKGGNMSSNG	ARAKGATGGS	1080
GGCCNGGMNN	MGMGGANMGA	GAJNGMGMB	GGGGGKTHK	TKKCTCCGNS	CGANGRAGAA	1140
GKTGNGSCCT	CGMGKYGRT	KTKTKTKTS	CTCTGCTMMN	NAGAAAAGAG	AGGGC	1195

(2) INFORMATION FOR SEQ ID NO:342:

(1) SEQUENCE CHARACTERISTICS:

(11) MOLECULE TYPE: Genomic DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:342

CCATCTGATC	GTGCGCAACC	AGCATCGCAG	TGGGAACGAT	GCCCTCATTC	AGCATTTCGA	60
TGGTTTGTG	AAAACCGGAC	ATGGCACTCC	AGTCGCCCTC	CCGTTCCGCT	ATCGGCTGAA	120
TTTGATTGCG	AGTGACATAT	TTATGCCAGC	CAGCCAGACG	CAGACGCGCC	GAGACAGAAC	180
TTAATGGGCC	CGCTAACAGC	GCGATTTGCT	GGTGACCCAA	TGCGACCAGA	TGCTCCACGC	240
CCAGTCGCGT	ACCGTCTTCA	TGGGAGAAAA	TAATACTGTT	GATGGGTGTC	TGGTCAGAGA	300
CATCAAGAAA	TAACGCCGGA	ACATTAGTGC	AGGCAGCTTC	CACAGCAATG	GCATCCTGGT	360
CATCCAGCGG	ATAGTTAATG	ATCAGCCAC	TGACGCGTTG	CGCGAGAAGA	TTGTGCACCG	420
CCGCTTTACA	GGCTTCGACG	CCGCTTCGTT	CTACCATCSA	CACCACCACG	CTGGCACCCA	480
GTTGATCGGC	GCGAGATTTA	ATCSCCCGSA	CAATTTGCGA	CGGCGCGTGC	AGGGCCAGAC	540
TGGAGGTGGC	AACGCCAATC	AGCAACGACT	GTTTGCCCCG	CAGTTGTTGT	GCCACGCGGT	600
TGGGAATGTA	ATTGAGCTCC	GCCATCGCCG	CTTCGACTTT	TTCCCGCGTT	TTCCGAGAAA	660
CGTGGCTGGC	CTGGTTCACC	ACGCGGGA	CGGTCTGATA	AGAGACACCG	GCATACCTTG	720
CGACATGCTA	TAACCTTACT	GGTTTCACAT	TCACCCACCT	GAATTGACTC	TCTTCCGGGC	780
CTATGATGTC	CATACCGCGA	AAGGTTTTGC	GCGATTCGAT	GCTGTCGGGG	ATCTCGACGC	840
TCTGCTTTAT	CGGACTCCTG	CATTAGGAAG	CAGTCCAGTA	GTAGGTTGAG	GCCGTTGAGC	900
ACCGCGCGCG	CAAGGAATGG	TGCATGCAAG	GAGATGGCGC	CCAACAGTCC	CCCGGCCACG	960
GGGCTTGCCA	CCATACCCAC	GCCGAAACAA	GCGTTCATGA	GCCCGAAGTG	GCGAGCGTGA	1020
TCTTCCCAT	CGGTGATGTC	GSCGATATAG	GCGCCAGCAA	CCGCACCTGT	GGCGCGCGTG	1080
ATGCGGGCCA	CGATGCCGTC	GCGGTAGAGG	ATCGAGATCT	CGATCCCGCG	AAATTAATAC	1140
GACTCACTAT	AGGGGAATTC	TGAGCGGATA	ACAAATCCCC	TCTAGAAATA	ATTTTGTGTA	1200
ACTTTAAGAA	GGAGATATAC	ATATGGGCGA	TCATCATCAT	CATCAGCTGA	TGCACATCAT	1260
CGGACCGAGC	CCCACATCCT	GCGAACAGGC	GGCGSCGAG	GCGGTCCAGC	GGGCGCGGGA	1320
TACCGTGGAT	GACATCGCGC	TGCGTCCGCT	CATTGAGCAG	GACATGGCGC	TGGACAGCCG	1380
CGGCAAGATC	ACCTACCGCA	TCAAGCTCGA	AGTGTGCTTC	AAGATGAGGC	CGGCGCAACC	1440
GATGGGCTCG	AAACGACCGA	GCGGTTCCGC	TGAACCGGCG	GCCGCGCGCG	GTACTGTGCG	1500
GACTACCGCC	GCGTGGTCCG	CGGTGACCTT	GCGGAGAGCC	GCTAGCACCG	TGCTGTACCG	1560
GCTGTTCAAC	CTGTGGGGTC	CGGCTTTTGA	CGAGAGGTAT	CGAACGTCA	CGATCAGCGG	1620
TCAGGGCACC	CGTTCTGGTG	CGGGGATCGC	CGAGGCGGCG	GCGGGGACCG	TCAACATTCG	1680
GCGCTCGGAC	GCTATCTGTT	TGGAAAGTGA	TATCGCGCGG	TACAAGGGGC	TGATGAACAT	1740
GCGGTAGCG	ATCTCGGCTC	AGCAGGTCGA	GTACAACTTG	TCCGGAGTGA	CGAAGGACCT	1800
CAAGGTGAAC	CGAAAAGTCC	TGCGCGGCAT	GTACAGGGCG	ACCATCAAAA	CGTGGGACGA	1860
CGCGGAGATC	GCTGCGCTCA	ACCGCGGCGT	GAACCTGCGC	GCGACCGCGG	TAGTTGCGCT	1920
CGAGGCTTCC	GACGGGTCGG	GTGACAGCTT	CTGTTCGAGC	CAGTACCTGT	CGAAGCAAGA	1980
TGCGGAGGGC	TGGGGCAAGT	CGGCGGCGCT	CGGCGCGAGC	CTCGACTTCC	CGGCGGTGCG	2040
TGCTGCGCTG	GCTGAGAACG	TCAATGCGCG	CTGTGTCAGC	GCTTTCGCGC	AGACAGCTTG	2100
TTGCTGCGCT	TATATCGGCA	TCAAGCTGCT	TGAGGAGGCG	AGTCAGCGCG	GACTGCGCTA	2160
GGGCAACTA	GGCAATAGCT	ATGGCAATTT	CTGTGTCAGC	GAJGGGCAAA	GCATTTAGAT	2220
CTGCTGCGCT	GGATTGCGAT	TGAAAATCTT	CTGTGTCAGC	CTGTGTCAGC	TGATCGAGCT	2280
CTGCTGCGCG	GACGCTAGC	CGATGATGAA	CTGTGTCAGC	CTGTGTCAGC	AGAACCGGTA	2340
GAAGGACGCG	GCGACCGCGC	AGAGCTTTCG	CTGTGTCAGC	CTGTGTCAGC	TCACCGACCG	2400
GAAGGAGCGC	TGCTTCTCTG	ACGAGGTTCA	CTGTGTCAGC	CTGTGTCAGC	CGGTGCTGAA	2460
CTGTGTCAGC	CGGTGTCAGC	CGACGATTTG	CTGTGTCAGC	CTGTGTCAGC	ATGAAAGACG	2520
CTGTGTCAGC	GAGGCAGGTA	ATTTGAGGCG	CTGTGTCAGC	CTGTGTCAGC	CGCAGATCGA	2580
CTGTGTCAGC	TCGACGGCAG	CTGTGTCAGC	CTGTGTCAGC	CTGTGTCAGC	CGGCGACCGT	2640
CTGTGTCAGC	CGGCTGCTGC	CTGTGTCAGC	CTGTGTCAGC	CTGTGTCAGC	AAAGGAGAAC	2700
CTGTGTCAGC	AGCAATATTC	CTGTGTCAGC	CTGTGTCAGC	CTGTGTCAGC	AGGAGATCGA	2760
CTGTGTCAGC	CTGTGTCAGC	CTGTGTCAGC	CTGTGTCAGC	CTGTGTCAGC	AGGAGATCGA	2820

```

AACCGTCTCG CTYGACGCCA ACGGGGTGTC TGGAAAGCGG TCGTATTACG AAGTCAAGTT 3120
CAGCGATCCG AGTAAGCCGA ACGGCCAGAT CTGGACGGGC GTAATCGGCT CGCCCGCGGC 3180
GAACGCACCG GACGCCGGGC CCCCTCAGCG CTGGTTTGTC GTATGGCTCG GGACCGCCAA 3240
CAACCCGGTG GACAAGGGCG CGGCCAAGGC GCTGGCCGAA TCGATCCGGC CTTTGGTGCG 3300
CCCGCCGGCG GCGCCGGGCG GGGAAAGTCG TCCTACCCCG ACGACACCGA CACCGCAGCG 3360
GACCTTACCG GCCTGAGAAT TCTGCAGATA TCCATCACAC TGGCGGGCCG TCGAGCACCA 3420
CCACCACCAC CACTGAGATC CGGCTGCTAA CAAAGCCCCG AAGGAAGCTG AGTTGGCTGC 3480
TGCCACCGCT GAGCAATAAC TAGCATAACC CCTTGGGGCC TCTAAACGGG TCTTGAGGGG 3540
TTTITTGCTG AAAGGAGGAA CTATATCCGG AT 3572

```

(2) INFORMATION FOR SEQ ID NO:343:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:343:

```

Val Gln Phe Gln Ser Gly Gly Asp Asn Ser Pro Ala Val Tyr Xaa Xaa
 1             5             10             15
Asp Gly Xaa Arg
          20

```

(2) INFORMATION FOR SEQ ID NO:344:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:344:

```

Thr Thr Val Phe Val Val Thr Val Ala Arg
 1             10

```

(2) INFORMATION FOR SEQ ID NO:345:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

1 S 10

(2) INFORMATION FOR SEQ ID NO: 346:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:346:

Asp Ala Gly Lys Xaa Ala Gly Xaa Asp Val Xaa Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO:347:

(2) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: peptide

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:347:

Thr Xaa Glu Glu Xaa Glu Glu Ser Phe Asn Ser Ala Ala Pro Gly Asn
1 5 10 15
Xaa Lys

2 INFORMATION FOR SEQ ID NO: 348:

1. SEQUENCE CHARACTERISTICS

- A. LENGTH: 27 base pairs
B. TYPE: nucleic acid
C. STRANDEDNESS: single
D. TOPOLOGY: linear

17 MOLECULE TYPE: Other

XREF SEQUENCE DESCRIPTION, SEQ ID NO. 348

THGTTAGTA TTCAGTCCA JACCGTG

REF ID: A66084

— 207 —

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 349:

GCAGTGACGA ATTCACCTCG ACTCC

25

(2) INFORMATION FOR SEQ ID NO:350:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2412 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

XL1 SEQUENCE DESCRIPTION: SEQ ID NO:350:

CATATGGGGCG	ATCATCATCA	TCATCAAGTG	ATCGACATCA	TGGGGACCAG	CCCCACATCG	50
TGGGAACAGG	CGGCGGCGGA	GGGGGTCCAG	CGGCGCGGGG	ATAGCCTCGA	TGACATCCGC	120
GTCGGTCGGG	TCATTGAGCA	GGATATGGCG	GTGGACAGCG	CGGCGAAGAT	CACCTACCGC	180
ATCAAAGCTCG	AAGTGTCTGT	CAAGATAGG	CGGCGCGAAG	CGAGGGGCTC	GAACACACCG	240
AGCGGTTCGT	CTGAAACGGG	CGCGGGCGCG	GCTACTGTCT	CGACTACCGC	CGCGTCTCTG	300
CGCGGTACCT	TGGCGGAGAG	CGGTAGCAAG	CTGCTCTACC	CGCTCTTCAA	CGTGTGCGGT	360
CGCGCGTTTG	ACGAGAGGTA	TCCGGAAGTG	ACGATCAAGC	CTCAGGGGCA	CGGTTCTGGT	420
CGCGCGATCG	CGCAGGCGCG	CGCGCGAGCG	CTCAAAATTG	GGCGCTCGAG	CGCTATCTG	480
TCGGAAGGTG	ATATGGCGCG	GCACAAAGGG	GTGATGAACA	TGGCGCTAGC	CATCTCGGCT	540
CAGCAAGTCA	ACTACAAGCT	CGCGGAGGTG	AGCGAGCAAC	TCAGGTGAA	CGGAAAAAGT	600
CTGCGCGGCA	TGTACCAAGG	TACCATCAAA	ACCTGGGAGC	ACCTCGAGAT	TGCTGCGCTG	660
AACTCGCGCG	TGAACCTCGC	CGGCAACCGG	TGAGTTCTCG	TGCAAGGCTT	CGACGGGCTC	720
GCTGACAGCT	TGTTGTTCACT	CGAGTACCTG	TGCAAGCAAG	ATCGGTAGGG	TGGGGCGAAG	780
TGCGCGCGCT	TGCGCACCACT	CGTCACTTC	CGCGCGGTCT	CGGCTGCGCT	TGGTGAAGAC	840
TGCAACCGCG	CGATGCTGAG	CGGTTGCGCG	GAGACAAAGG	CGTGGCTTGG	TTATATCGG	900
ATCAAGTTTG	TGACCAAGCG	CGGTCACCG	CGACTCGGG	AGGCGCAACT	AGGCAATAGC	960
TGTGCGAATT	TGTTGTTTCC	CGAGCGCGAA	AGCATTTAGG	CGCGCGCGCG	TGCTCTGCA	1020
TGAAAAAGCG	CGCGCAAGCA	GGCGATTTCG	ATGATCGAG	GGCGCGCGCG	GGACCGGTAC	1080
CGATATATCA	ACTACAGGTA	CGGCATCTCT	AAGCAACCGC	AAAAAGAGCG	CGCGACGCG	1140
TAGACCTTTC	AGGCATTCTT	CGATCTGGCG	ATCAAGAGCG	CGAAAGAGCG	CTGCTTCTCT	1200
TAGCAAGTTT	ATTTCTAGCG	CGTGGCGCGG	CGGCTGCTGA	AGTTCTCTGA	TGCTTGGATT	1260
TGCAAGATTG	CGAGCGCTCA	CATCAAGAGC	TATCGCGCTA	CGCTCGCGCA	CGAGGCGAGG	1320
AGTTTCTGAG	CGATCTCGCG	CGAGCTGAAA	AGCGAGATCG	AGAGGTGCA	TTGATCGCA	1380
CGTTCTGCTG	AGGCGAGTGG	CGCGCGCGCG	CGCGCGAGCG	CGCGCGAGCG	CGCGGTGGTG	1440
CGTTTCTAGG	AAGCAAGCA	TAGCAGAGAA	CAGGAACTCG	AGGATCTCT	GAGCAATATT	1500
CGTCAAGGCG	CGCTCGAATA	CTCGAGGCGG	GACGAGGAGG	AGCAGCGAGG	CGCTCTCTCT	1560
CAATATGGCT	TTGTGCGCAC	AAGCGCGCGG	TGCGCGCTCT	CGACTCTCTG	AGCGCGCGCG	1620
CGACGGGCGA	CAGCTGTTGG	CGCGCGACCA	CGGCGCGCGG	CGACACCGCG	GAATCGCGAG	1680
CGGCGCGATG	CGAAGCGAGC	ACCTCGCGCG	CGCGACCGCA	AGCGACCGCG	GCACCTCTCT	1740
ATTTCTGAGG	AAGCAAGCA	ACCTCGCGCG	ATGCAAGAGC	CGCTTGGAGG	ATTCAGCTTT	1800
CGCTCTGCTG	CGGCTGAGGT	CGAGTTGAG	CGCGCTACT	TGCACTAGCG	TTGAGCACTT	1860

```

ATCGGCTCGC CCGCGGCGAA CGCACCGGAC GCGGGGCCCC CTCAGCGCTG GTTTGTGGTA 2220
TGGCTCGGGA CCGCCAACAA CCGGVTGGAC AAGGGCGCGG CCAAGGCGCT GGCCGAATCG 2280
ATCCGGCCTT TGGTCGCCCC GCGGCGGGCG CCGGCACCGG CTCCTGCAGA GCGCGCTCCG 2340
GCGCCGCGCG CCGCCGGGGA AGTCGCTCCT ACCCCGACGA CACCGACACC GCAGCGGACC 2400
TTACCGGCCT GA 2412

```

(2) INFORMATION FOR SEQ ID NO:351:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 802 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:351:

```

Met Gly His His His His His His Val Ile Asp Ile Ile Gly Thr Ser
  1           5           10           15
Pro Thr Ser Trp Glu Gln Ala Ala Ala Glu Ala Val Gln Arg Ala Arg
  20           25           30
Asp Ser Val Asp Asp Ile Arg Val Ala Arg Val Ile Glu Gln Asp Met
  35           40           45
Ala Val Asp Ser Ala Gly Lys Ile Thr Tyr Arg Ile Lys Leu Glu Val
  50           55           60
Ser Phe Lys Met Arg Pro Ala Gln Pro Arg Gly Ser Lys Pro Pro Ser
  65           70           75           80
Gly Ser Pro Gln Thr Gly Ala Gly Ala Gly Thr Val Ala Thr Thr Pro
  85           90           95
Ala Ser Ser Pro Val Thr Leu Ala Gln Thr Gly Ser Thr Leu Leu Tyr
 100           105           110
Pro Leu Phe Asn Leu Trp Gly Pro Ala Phe His Glu Arg Tyr Pro Asn
 115           120           125
Val Thr Ile Thr Ala Gln Gly Thr Gly Ser Gly Ala Gly Ile Ala Gln
 130           135           140
Ala Ala Ala Gly Thr Val Asn Ile Gly Ala Ser Asp Ala Tyr Leu Ser
 145           150           155           160
His Gly Asp Met Ala Ala His Lys Gly Leu Met Asn Ile Ala Leu Ala
 165           170           175
Ile Ser Ala His Gln Val Asn Tyr Asn Leu Pro Gly Val Ser Gln His
 180           185           190
Leu Lys Leu Asn Gly His Val Leu Ala Ala Met Thr Gln Gly Thr Ile
 195           200           205
Lys Thr Thr Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn
 210           215           220
Leu Pro Gly Thr Ala Val Val Pro Leu His Arg Ser Asn Gly Ser Gly
 225           230           235           240
Asp Thr Phe Leu Phe Thr Gln Tyr Leu Ser Leu His Asn Pro Gln Gly
 245           250           255

```

272

290 295 300
 Gln Ala Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser
 305 310 315 320
 Gly Asn Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala
 325 330 335
 Gly Phe Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp
 340 345 350
 Gly Pro Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile
 355 360 365
 Val Asn Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala
 370 375 380
 Phe Leu His Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp
 385 390 395 400
 Gln Val His Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp
 405 410 415
 Ala Leu Ile Ala Thr Ile Ser Ser Ala Glu Met Lys Thr Asp Ala Ala
 420 425 430
 Thr Leu Ala Gln Glu Ala Gly Asn Phe Glu Arg Ile Ser Gly Asp Leu
 435 440 445
 Lys Thr Gln Ile Asp Gln Val Glu Ser Thr Ala Gly Ser Leu Gln Gly
 450 455 460
 Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln Ala Ala Val Val Arg
 465 470 475 480
 Phe Gln Glu Ala Ala Asn Lys Gln Lys Gln Glu Leu Asp Glu Ile Ser
 485 490 495
 Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg Ala Asp Glu Glu
 500 505 510
 Gln Gln Gln Ala Leu Ser Ser Gln Met Gly Phe Val Pro Thr Thr Ala
 515 520 525
 Ala Ser Pro Pro Ser Thr Ala Ala Ala Pro Pro Ala Pro Ala Thr Pro
 530 535 540
 Val Ala Pro Pro Pro Pro Ala Ala Ala Asn Thr Pro Asn Ala Gln Pro
 545 550 555 560
 Gly Asp Pro Asn Ala Ala Pro Pro Pro Ala Asp Pro Asn Ala Pro Pro
 565 570 575
 Pro Pro Val Ile Ala Pro Asn Ala Pro Gln Pro Val Arg Ile Asp Asn
 580 585 590
 Pro Val Gly Gly Phe Ser Phe Ala Leu Pro Ala Gly Trp Val Glu Ser
 595 600 605
 Asn Ala Ala His Phe Asp Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr
 610 615 620
 Gly Asp Pro His Phe Leu Gly Gln Pro Ser Pro Val Ala Asn Asp Thr
 625 630 635 640
 Arg Ile Val Leu His Arg Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu
 645 650 655
 Ala Thr Asp Ser Lys Ala Ala Ala Arg Leu Gly Ser Asp Met Gly Glu
 660 665 670
 Phe Tyr Met Pro Tyr Pro Gly Thr Arg Ile Asn Gln Glu Thr Val Ser
 675 680 685
 Leu Asn Ala Asn Gly Val Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys

Phe Val Val Trp Leu Gly Thr Ala Asn Asn Pro Val Asp Lys Gly Ala
 740 745 750
 Ala Lys Ala Leu Ala Glu Ser Ile Arg Pro Leu Val Ala Pro Pro Pro
 755 760 765
 Ala Pro Ala Pro Ala Pro Ala Glu Pro Ala Pro Ala Pro Ala Pro Ala
 770 775 780
 Gly Glu Val Ala Pro Thr Pro Thr Thr Pro Thr Pro Gln Arg Thr Leu
 785 790 795 800
 Pro Ala

(2) INFORMATION FOR SEQ ID NO:352:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Other

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:352:

GGATCCAAAC CACCGAGCGG TTCGCCTGAA ACGG

34

(2) INFORMATION FOR SEQ ID NO:353:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: Other

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:353:

CGCTGTGAAT TCACCTCCGG AGGAAATCGT CCGCATC

37

(2) INFORMATION FOR SEQ ID NO:354:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1962 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(11) MOLECULE TYPE: cDNA

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:354:

```

GCCGCCGGGA CGGTCAACAT TGGGGCCTTC GACGCCTATC TGTCGGGAAGG TGATATGGCC 300
GCGCACAAAG GGCTGATGAA CATCGCCTTA GCCATCTCCG CTCAGCAGGT CAACTACAAC 360
CTGCCCGGAG TGAGCGAGCA CCTCAAGCTG AACGGAAAAG TCCTGGCGGC CATGTACCAG 420
GGCACCATCA AAACCTGGGA CGACCCGCGAG ATCGCTGCGC TCAACCCCGG CGTGAACCTG 480
CCCGGCACCG CGGTAGTTCC GCTGCACCGC TCCGACGGGT CCGGTGACAC CTTCTTGTTT 540
ACCCAGTACC TGTCCAAGCA AGATCCCGAG GGCTGGGGCA AGTCGCCCGG CTTCCGGCACC 600
ACCGTCCGACT TCCCGGCGGT GCCGGGTGCG CTGGGTGAGA ACGGCAACGG CGGCATGGTG 660
ACCGGTTGCG CCGAGACACC GGGCTGCGTG GCCTATATCG GCATCAGCTT CCTCGACCAG 720
GCCAGTCAAA GGGGACTCGG CGAGGCCCAA CTAGGCAATA GCTCTGGCAA TTTCTTGTTG 780
CCCGACGCGC AAAGCATTCA GGCCGCGGCG GCTGGCTTCG CATCGAAAAC CCCGCGCAAC 840
CAGGCGATTT CGATGATCGA CGGGCCCGCC CCGGACGGCT ACCCGATCAT CAACTACGAG 900
TACGCCATCG TCAACAACCG GCAAAGGAC GCCGCCACCG CGCAGACCTT GCAGGCATTT 960
CTGCACTGGG CGATCACCAG CCGCAACAAG GCCTCGTTCC TCGACCAGGT TCAITTCAG 1020
CCGCTGCCCG CCGCGGTGGT GAAGTTGTCT GACGCGTTGA TCGCGACGAT TTCTCCGGA 1080
GGTGGCAGTG GGGGAGGCTC AGGTGGAGGT TCTGGCGGGA GCGTGCCCGC AACGGCCGCC 1140
TCGCCCGCGT CGACCGCTGC AGCGCCACCC GCACCGGCGA CACCTGTTGC CCCCCACCA 1200
TGGGCGCGCG CCGACCGCGC GAATGCGGAG CCGGGCGGAT CCAACGCAGC ACCTCGCGCG 1260
GCCGACCGCA ACGCACCGCG GCCACCTCTC ATTGCCCCAA ACGCACCGCA ACCTGTCGCG 1320
ATCGACAACC CGGTGGAGG ATTCAJCTTC GCGCTGCGTG CTGGCTGGGT GGAGTCTGAC 1380
GCCGCCCACT TCGACTACCG TTCAGCACTC CTCAGCAAAA CCACCGGGGA CCCGCCATTT 1440
CCCGGACAGC CGCCGCGGCT GGCCAATBAC ACCCGTATCG TGCTCGGCGG GCTAGACCAA 1500
AAGCTTTAGC CCAGCGCGGA AGCCACCGAC TCCAAGGCGG CGGCGCGGTT GGCTCGGAC 1560
ATGGGTGAGT TCTATATGCC CTACCCGGGT ACCCGGATCA ACCAGGAAAC CGTCTCGCTC 1620
GACGTCACAG GGGTGTCTGG AAGCGCGTCT TATTACGAAG TCAAGTTTCA CGATCCGAGT 1680
AAACCGAAGC GCCAGATCTG GACGGCGCTA ATCGGCTCGC CCGCGGCGAA CGCACCGGAC 1740
GCCGCGCGCG CTCAGCGGTG TTTTGTGCTA TGGCTCGGGA CCGCCAACAA CCGGTGGGAC 1800
AAGGCGCGCG CCAAGGCGCT GCGGCAATCG ATCGCGCCTT TGGTCCGCGG GCGCGCGGCG 1860
CCCGIACCGG CTCCTGCAGA GCGCGCTCTG GCGCGGCGCG CGGCGGGGA AGTCGCTCCT 1920
ACCCGACCGA CACCGACACC GCAGCGGACT TTACCGGCTT GA 1960

```

2. INFORMATION FOR SEQ ID NO:55:

1. SEQUENCE CHARACTERISTICS:

- A. LENGTH: 552 amino acids
- B. TYPE: amino acid
- C. STRANDEDNESS: single
- D. TOPOLOGY: linear

11. MOLECULE TYPE: protein

11. SEQUENCE DESCRIPTION: SEQ ID NO:55

```

Met Val His Ala Val His His His Gly Ser Gly Pro Pro Ser Gly Ser
1 5 10 15
Pro Val Thr Gly Ala Gly Ala Gly Thr Val Ala Thr Thr Pro Ala Ser
20 25 30
Ser Thr Val Thr Leu Ala Gly Thr Gly Ser Thr Leu Leu Tyr Pro Leu
35 40 45
Pro Ala Leu Thr Gly Pro Ala Pro His His Ala Thr Pro Ala Val Thr
50 55 60
Thr Thr Ala His His Thr
65

```


275

100 105 110
 Ala Gln Gln Val Asn Tyr Asn Leu Pro Gly Val Ser Glu His Leu Lys
 115 120 125
 Leu Asn Gly Lys Val Leu Ala Ala Met Tyr Gln Gly Thr Ile Lys Thr
 130 135 140
 Trp Asp Asp Pro Gln Ile Ala Ala Leu Asn Pro Gly Val Asn Leu Pro
 145 150 155 160
 Gly Thr Ala Val Val Pro Leu His Arg Ser Asp Gly Ser Gly Asp Thr
 165 170 175
 Phe Leu Phe Thr Gln Tyr Leu Ser Lys Gln Asp Pro Glu Gly Trp Gly
 180 185 190
 Lys Ser Pro Gly Phe Gly Thr Thr Val Asp Phe Pro Ala Val Pro Gly
 195 200 205
 Ala Leu Gly Glu Asn Gly Asn Gly Gly Met Val Thr Gly Cys Ala Glu
 210 215 220
 Thr Pro Gly Cys Val Ala Tyr Ile Gly Ile Ser Phe Leu Asp Gln Ala
 225 230 235 240
 Ser Gln Arg Gly Leu Gly Glu Ala Gln Leu Gly Asn Ser Ser Gly Asn
 245 250 255
 Phe Leu Leu Pro Asp Ala Gln Ser Ile Gln Ala Ala Ala Ala Gly Phe
 260 265 270
 Ala Ser Lys Thr Pro Ala Asn Gln Ala Ile Ser Met Ile Asp Gly Pro
 275 280 285
 Ala Pro Asp Gly Tyr Pro Ile Ile Asn Tyr Glu Tyr Ala Ile Val Asn
 290 295 300
 Asn Arg Gln Lys Asp Ala Ala Thr Ala Gln Thr Leu Gln Ala Phe Leu
 305 310 315 320
 His Trp Ala Ile Thr Asp Gly Asn Lys Ala Ser Phe Leu Asp Gln Val
 325 330 335
 His Phe Gln Pro Leu Pro Pro Ala Val Val Lys Leu Ser Asp Ala Leu
 340 345 350
 Ile Ala Thr Ile Ser Ser Gly Gly Gly Ser Gly Gly Ser Gly Gly
 355 360 365
 Glu Ser Gly Gly Ser Val Pro Thr Thr Ala Ala Ser Pro Pro Ser Thr
 370 375 380
 Ala Ala Ala Pro Pro Ala Pro Ala Thr Pro Val Ala Pro Pro Pro Pro
 385 390 395 400
 Ala Ala Ala Asn Thr Pro Asn Ala Gln Pro Gly Asp Pro Asn Ala Ala
 405 410 415
 Ala Ala Ala Ala Asn Ala Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 420 425 430 435
 Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 440 445 450
 Leu Ala Leu Pro Ala Ala Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr
 455 460 465
 Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr Gly Asp Pro Pro Phe Pro
 465 470 475 480
 Gly Gln Pro Phe Phe Val Ala Asn Asp Thr Arg Ile Val Leu Gly Arg
 485 490 495
 Leu Asp Gln Lys Leu Thr Ala Ser Ala Ile Ala Thr Asn Ser Lys Ala
 500

Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys Phe Ser Asp Pro Ser Lys
 545 550 555 560
 Pro Asn Gly Gln Ile Trp Thr Gly Val Ile Gly Ser Pro Ala Ala Asn
 565 570 575
 Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp Phe Val Val Trp Leu Gly
 580 585 590
 Thr Ala Asn Asn Pro Val Asp Lys Gly Ala Ala Lys Ala Leu Ala Glu
 595 600 605
 Ser Ile Arg Pro Leu Val Ala Pro Pro Pro Ala Pro Ala Pro Ala Pro
 610 615 620
 Ala Glu Pro Ala Pro Ala Pro Ala Pro Ala Gly Glu Val Ala Pro Thr
 625 630 635 640
 Pro Thr Thr Pro Thr Pro Gln Arg Thr Leu Pro Ala
 645 650

CLAIMS

1. A polypeptide comprising an immunogenic portion of a soluble *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen has an N-terminal sequence selected from the group consisting of:

- (a) Asp-Pro-Val-Asp-Ala-Val-Ile-Asn-Thr-Thr-Cys-Asn-Tyr-Gly-Gln-Val-Val-Ala-Ala-Leu; (SEQ ID No. 120)
- (b) Ala-Val-Glu-Ser-Gly-Met-Ile-Ala-Leu-Gly-Thr-Pro-Ala-Pro-Ser; (SEQ ID No. 121)
- (c) Ala-Ala-Met-Lys-Pro-Arg-Thr-Gly-Asp-Gly-Pro-Leu-Glu-Ala-Ala-Lys-Glu-Gly-Arg; (SEQ ID No. 122)
- (d) Tyr-Tyr-Trp-Cys-Pro-Gly-Gln-Pro-Phe-Asp-Pro-Ala-Trp-Gly-Pro; (SEQ ID No. 123)
- (e) Asp-Ile-Gly-Ser-Glu-Ser-Thr-Glu-Asp-Gln-Gln-Xaa-Ala-Val; (SEQ ID No. 124)
- (f) Ala-Glu-Glu-Ser-Ile-Ser-Thr-Xaa-Glu-Xaa-Ile-Val-Pro; (SEQ ID No. 125)
- (g) Asp-Pro-Glu-Pro-Ala-Pro-Pro-Val-Pro-Thr-Thr-Ala-Ala-Ser-Pro-Pro-Ser; (SEQ ID No. 126)
- (h) Ala-Pro-Lys-Thr-Tyr-Xaa-Glu-Glu-Leu-Lys-Gly-Thr-Asp-Thr-Gly; (SEQ ID No. 127)
- (i) Asp-Pro-Ala-Ser-Ala-Pro-Asp-Val-Pro-Thr-Ala-Ala-Gln-Ileu-Thr-Ser-Ileu-Ileu-Asn-Ser-Ileu-Ala-Asp-Pro-Asn-Val-Ser-Phe-Xaa-Asn; (SEQ ID No. 128) and
- (j) Ala-Pro-Glu-Ser-Gly-Ala-Gly-Ileu-Gly-Gly-Thr-Val-Gln-Ala-Gly; (SEQ ID No. 130)

wherein Xaa may be any amino acid

substitutions and/or modifications, wherein said antigen has an N-terminal sequence selected from the group consisting of:

- (a) Asp-Pro-Pro-Asp-Pro-His-Gln-Xaa-Asp-Met-Thr-Lys-Gly-Tyr-Tyr-Pro-Gly-Gly-Arg-Arg-Xaa-Phe; (SEQ ID No. 129) and
- (b) Xaa-Tyr-Ile-Ala-Tyr-Xaa-Thr-Thr-Ala-Gly-Ile-Val-Pro-Gly-Lys-Ile-Asn-Val-His-Leu-Val; (SEQ ID No. 137), wherein Xaa may be any amino acid.

3. A polypeptide comprising an immunogenic portion of a soluble *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID Nos.: 1, 2, 4-10, 13-25, 52, 99 and 101, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 1, 2, 4-10, 13-25, 52, 99 and 101 or a complement thereof under moderately stringent conditions.

4. A polypeptide comprising an immunogenic portion of a *M. tuberculosis* antigen, or a variant of said antigen that differs only in conservative substitutions and/or modifications, wherein said antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID Nos.: 26-51, 138, 139, 163-183, 201, 240, 242-247, 253-256, 295-298, 309, 316, 318-320, 322, 324, 328, 329, 333, 335, 337, 339 and 341, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 26-51, 138, 139, 163-183, 201, 240, 242-247, 253-256, 295-298, 309, 316, 318-320, 322, 324, 328, 329, 333, 335, 337, 339 and 341 or a complement thereof under moderately stringent conditions.

5. A DNA molecule comprising a nucleotide sequence encoding a polypeptide according to any one of claims 1-4.

7. A host cell transformed with an expression vector according to claim 6.
8. The host cell of claim 7 wherein the host cell is selected from the group consisting of *E. coli*, yeast and mammalian cells.
9. A pharmaceutical composition comprising one or more polypeptides according to any one of claims 1-4 and a physiologically acceptable carrier.
10. A pharmaceutical composition comprising one or more DNA molecules according to claim 5 and a physiologically acceptable carrier.
11. A pharmaceutical composition comprising one or more DNA sequences recited in SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347; and a physiologically acceptable carrier.
12. A vaccine comprising one or more polypeptides according to any one of claims 1-4 and a non-specific immune response enhancer.
13. A vaccine comprising
a polypeptide having an N-terminal sequence selected from the group consisting of sequences recited in SEQ ID NOs: 134 and 135, and
a non-specific immune response enhancer
14. A vaccine comprising
one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215, 225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347, the complements of said sequences, and DNA sequences that

a non-specific immune response enhancer.

15. The vaccine of claims 12-14 wherein the non-specific immune response enhancer is an adjuvant.

16. A vaccine comprising one or more DNA molecules according to claim 5 and a non-specific immune response enhancer.

17. A vaccine comprising one or more DNA sequences recited in SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215, 225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347, and a non-specific immune response enhancer.

18. The vaccine of claims 16 or 17 wherein the non-specific immune response enhancer is an adjuvant.

19. A method for inducing protective immunity in a patient, comprising administering to a patient a pharmaceutical composition according to any one of claims 9-11.

20. A method for inducing protective immunity in a patient, comprising administering to a patient a vaccine according to any one of claims 12-18.

21. A fusion protein comprising two or more polypeptides according to any one of claims 1-4.

22. A fusion protein comprising one or more polypeptides according to any one of claims 1-4 and ESAT-6.

23. A fusion protein comprising one or more polypeptides according to any one of claims 1-4 and the *M. tuberculosis* antigen 18 kDa secretory protein.

24. A pharmaceutical composition comprising a fusion protein according to any one of claims 21-23 and a physiologically acceptable carrier.

25. A vaccine comprising a fusion protein according to any one of claims 21-23 and a non-specific immune response enhancer.

26. The vaccine of claim 25 wherein the non-specific immune response enhancer is an adjuvant.

27. A method for inducing protective immunity in a patient, comprising administering to a patient a pharmaceutical composition according to claim 24.

28. A method for inducing protective immunity in a patient, comprising administering to a patient a vaccine according to claims 25 or 26.

29. A method for detecting tuberculosis in a patient, comprising:

(a) contacting dermal cells of a patient with one or more polypeptides according to any one of claims 1-4; and

(b) detecting an immune response on the patient's skin and therefrom detecting tuberculosis in the patient.

30. A method for detecting tuberculosis in a patient, comprising:

(a) contacting dermal cells of a patient with a polypeptide having an N-terminal sequence selected from the group consisting of sequences recited in SEQ ID NO 134 and 135; and

(b) detecting an immune response on the patient's skin and therefrom detecting tuberculosis in the patient.

31. A method for detecting tuberculosis in a patient, comprising:

310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347; and

(b) detecting an immune response on the patient's skin and therefrom detecting tuberculosis in the patient.

32. The method of any one of claims 29-31 wherein the immune response is induration.

33. A diagnostic kit comprising:

(a) a polypeptide according to any one of claims 1-4; and
(b) apparatus sufficient to contact said polypeptide with the dermal cells of a patient.

34. A diagnostic kit comprising:

(a) a polypeptide having an N-terminal sequence selected from the group consisting of sequences recited in SEQ ID NO: 134 and 135; and
(b) apparatus sufficient to contact said polypeptide with the dermal cells of a patient.

35. A diagnostic kit comprising:

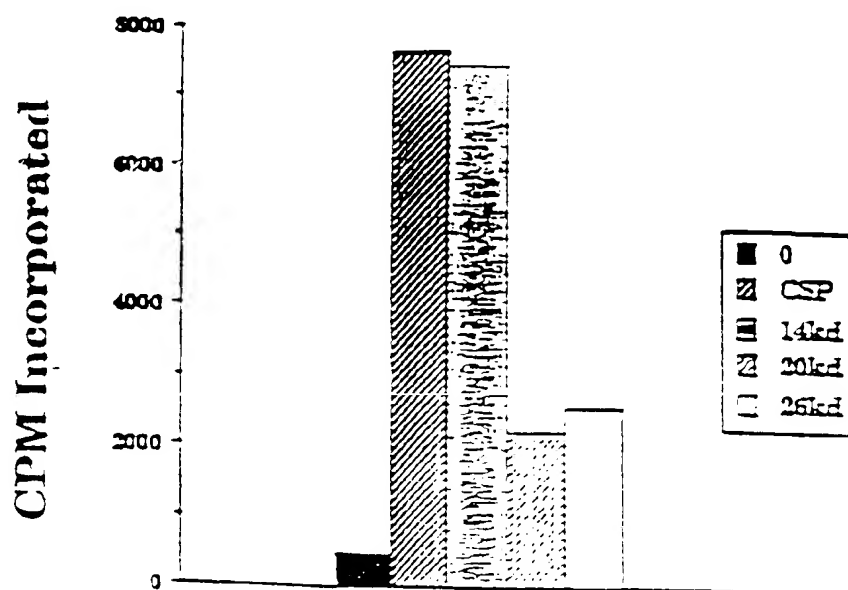
(a) a polypeptide encoded by a DNA sequence selected from the group consisting of SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID Nos.: 3, 11, 12, 140, 141, 156-160, 189-193, 199, 200, 203, 215-225, 237, 239, 261-276, 292, 293, 303-308, 310-315, 317, 321, 323, 325-327, 330-332, 334, 336, 338, 340 and 342-347; and

36. A diagnostic kit comprising:

- (a) a fusion protein according to any one of claims 21-23; and
- (b) apparatus sufficient to contact said fusion protein with the dermal cells of a patient.

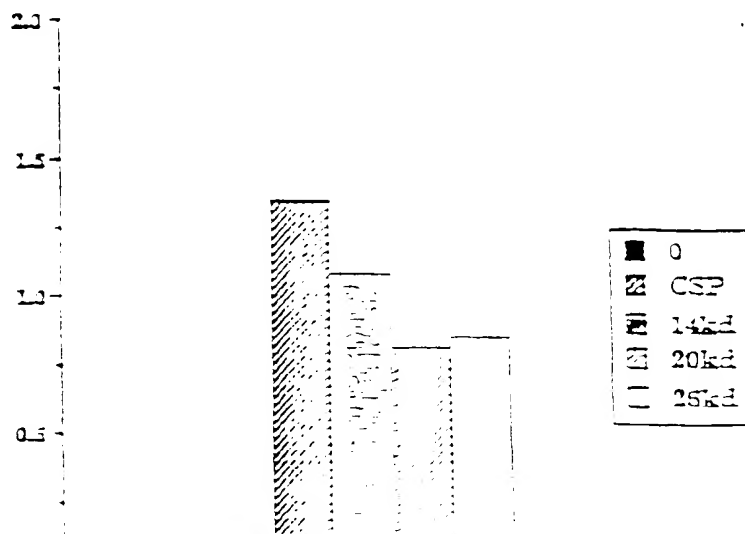
37. A fusion protein according to claim 23 comprising an amino acid sequence selected from the group consisting of sequences recited in SEQ ID NO: 153, 209, 351 and 355.

D7 T Cell Proliferation

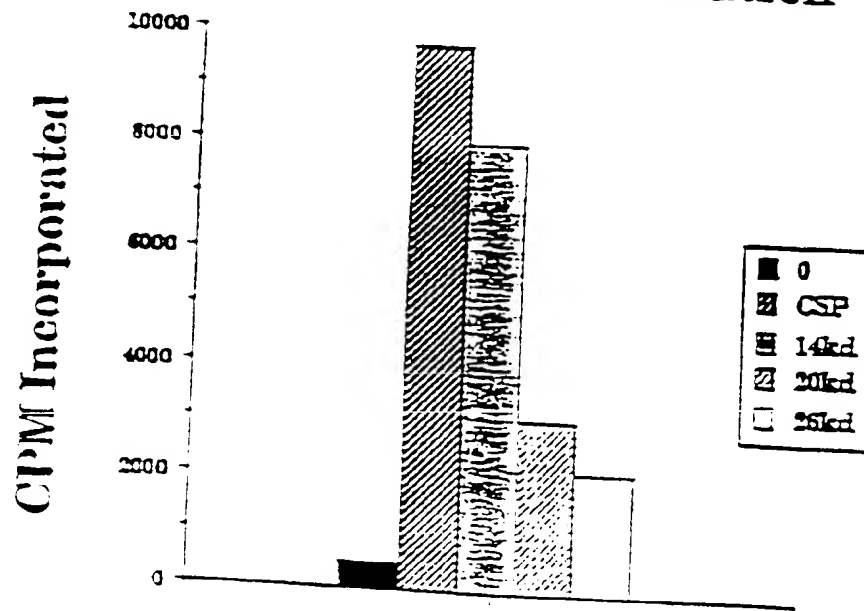


D7 IFN γ

O.D. 450-570

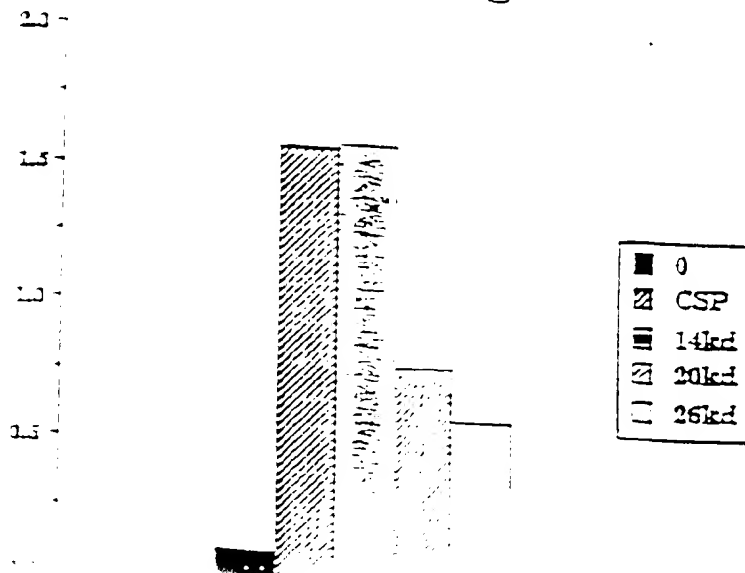


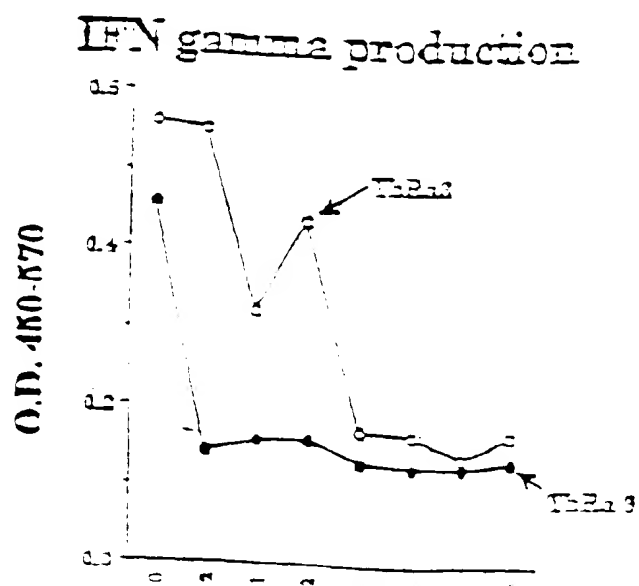
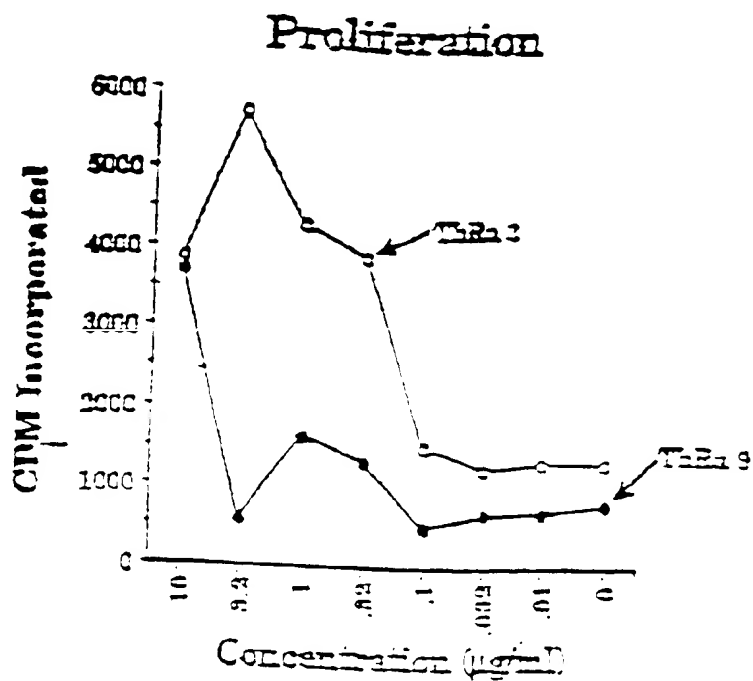
D160 T Cell Proliferation

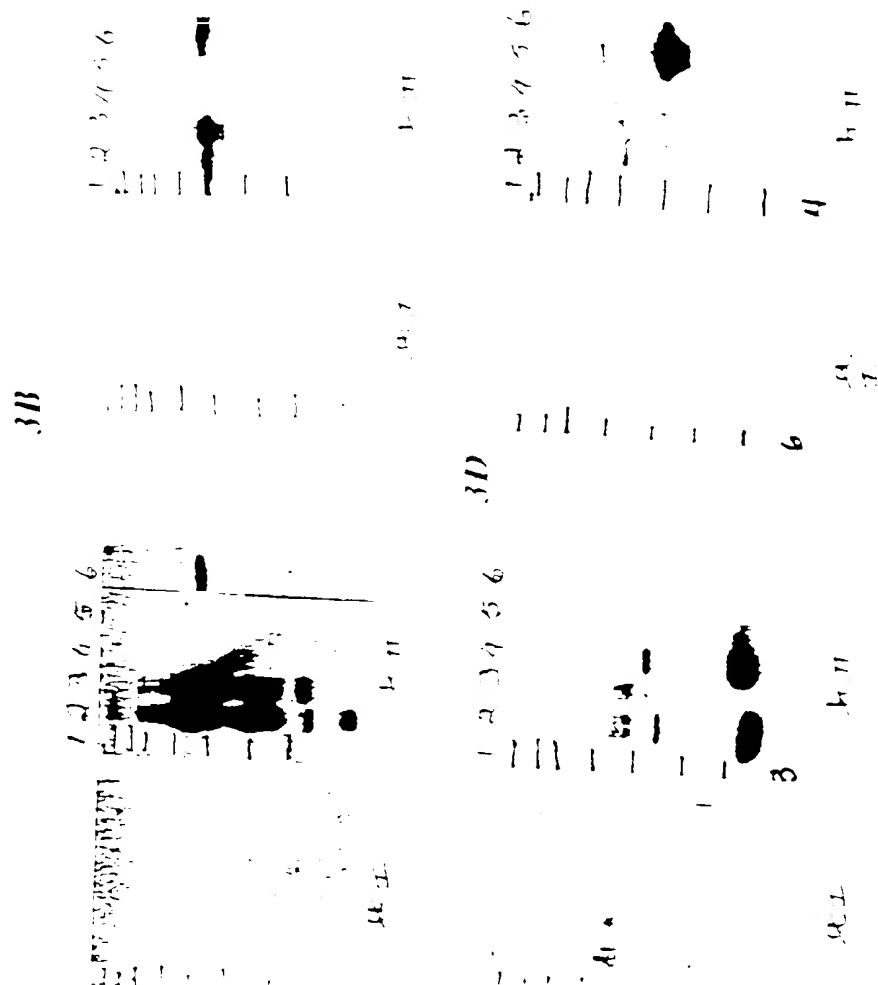


D160 IFN_g

O.D. 450-570







FIGS. 3A-D

T cell clone 131TbH9 responds poorly to CSP

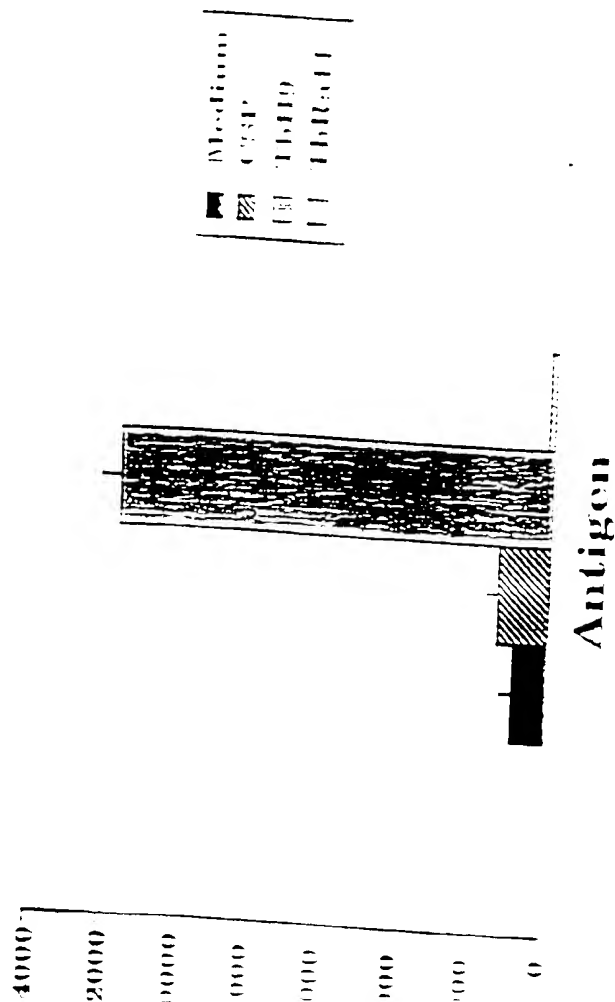


FIG. 4A

T Cell Clone PPD 800-10 IFN γ Production

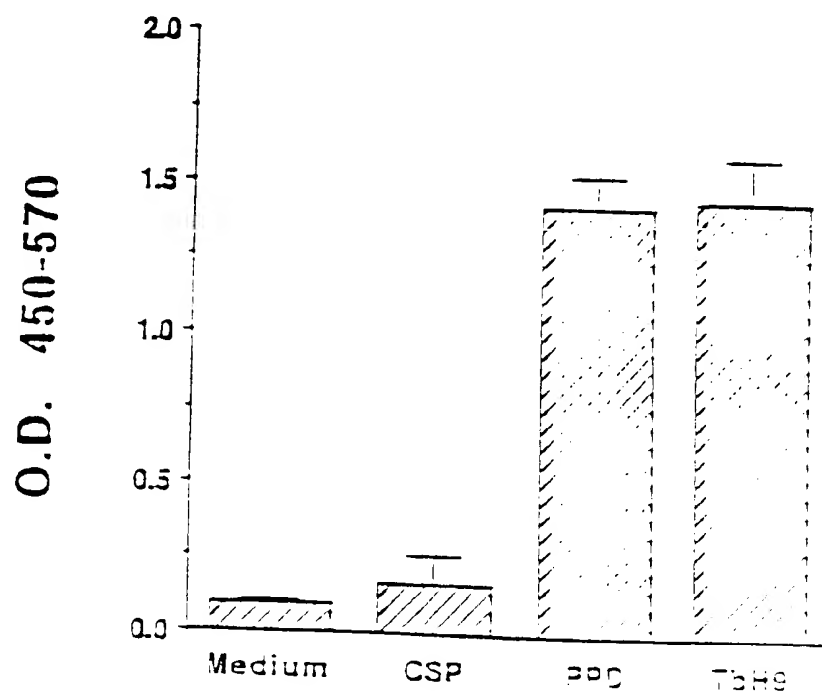
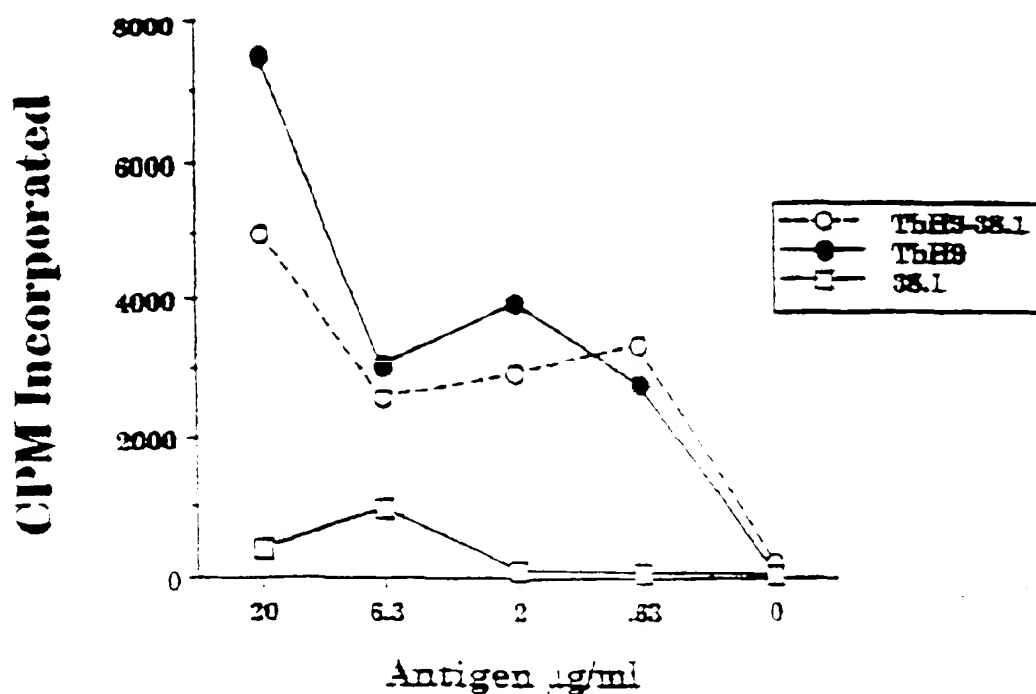
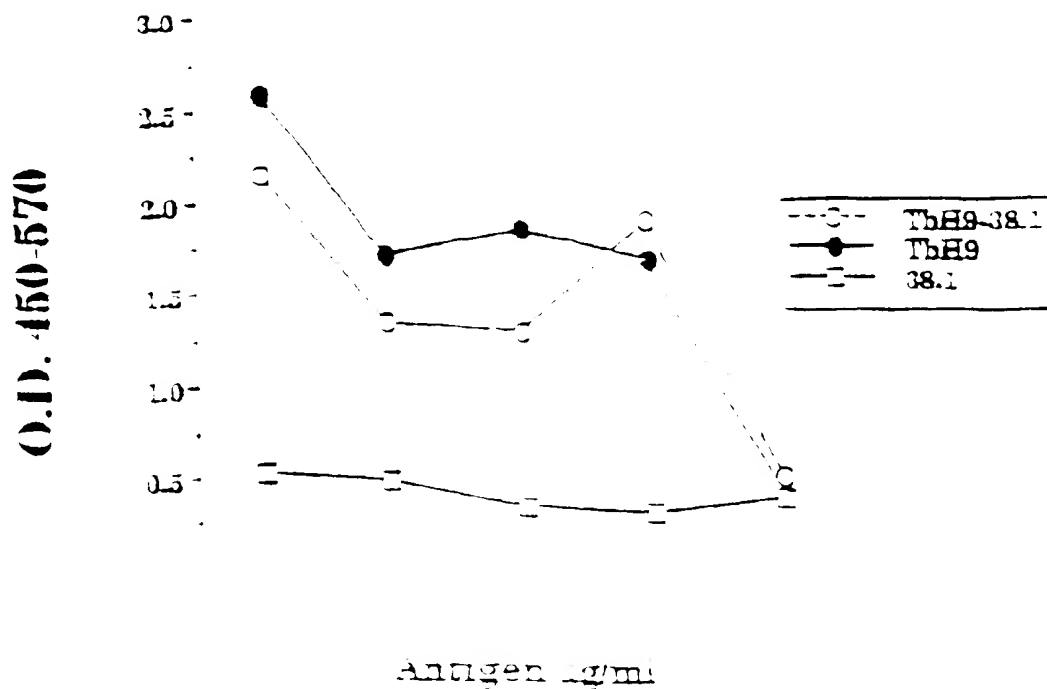


FIG. 4B

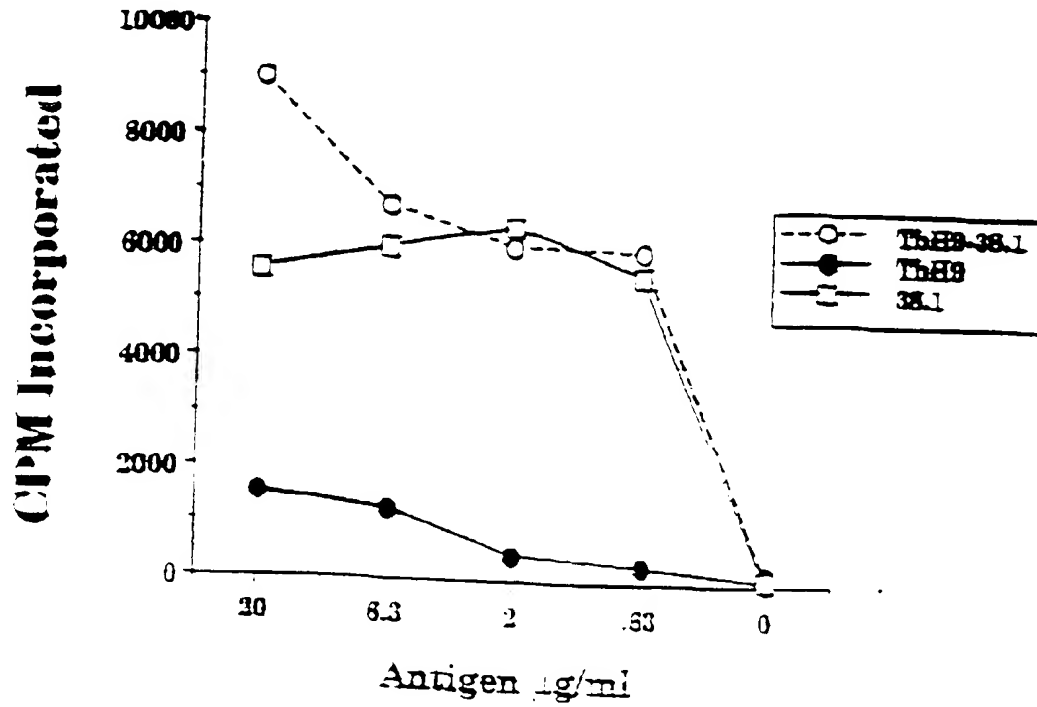
D131 T Cell Proliferation



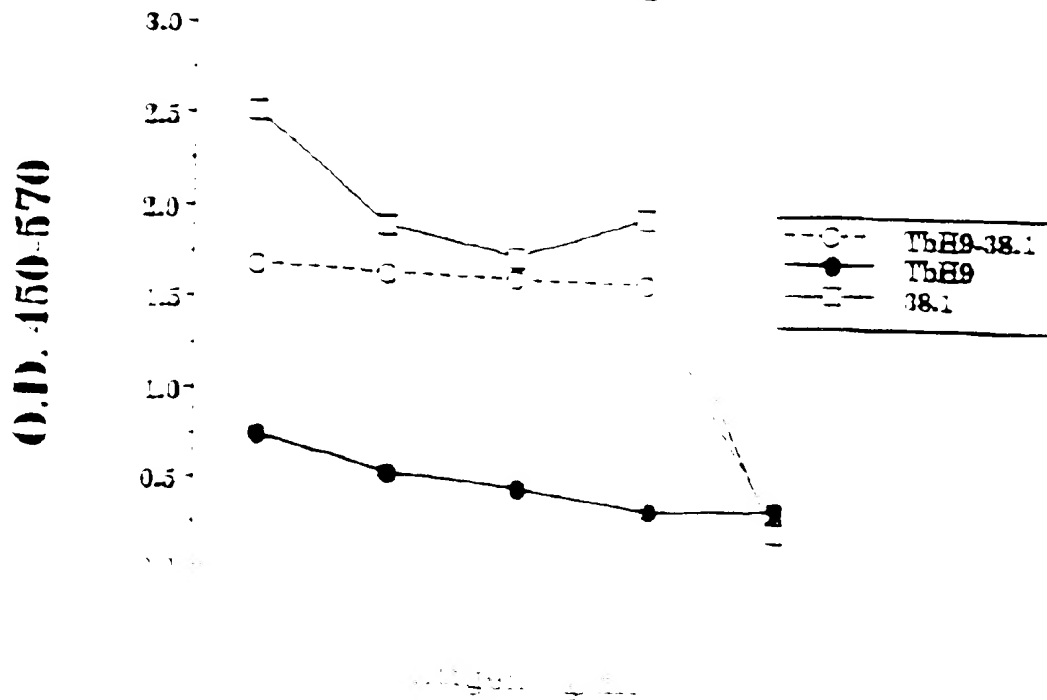
D131 IFN γ

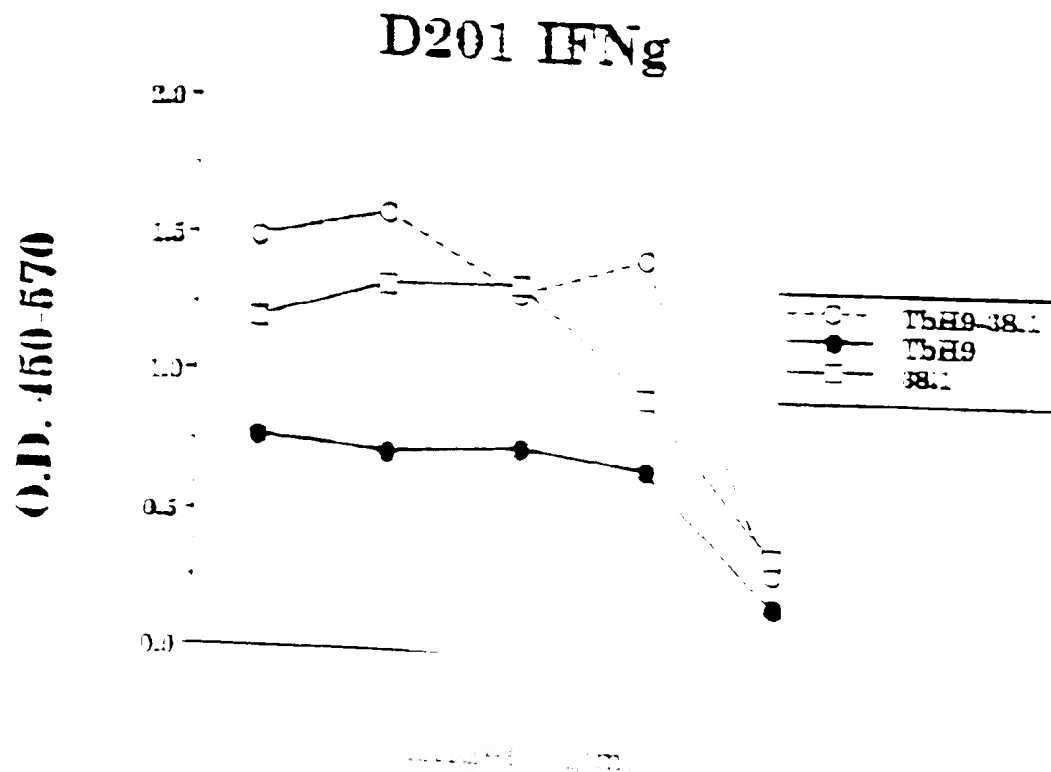
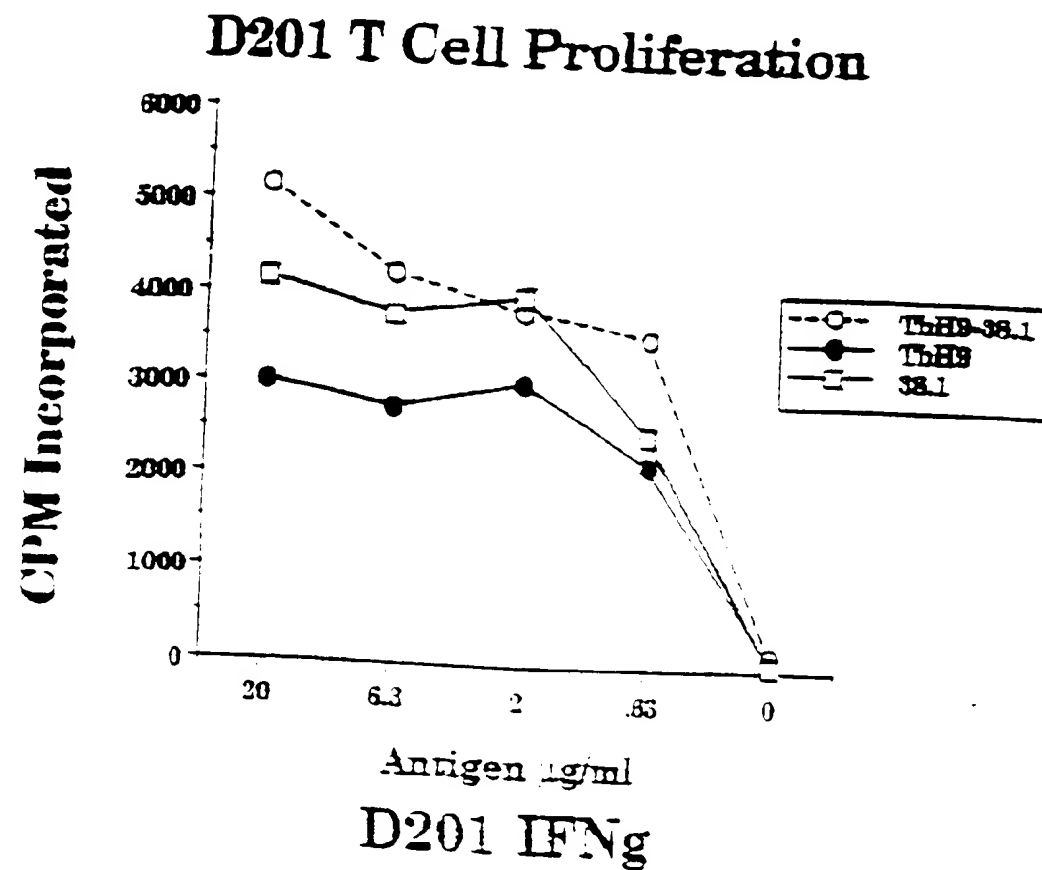


D184 T Cell Proliferation

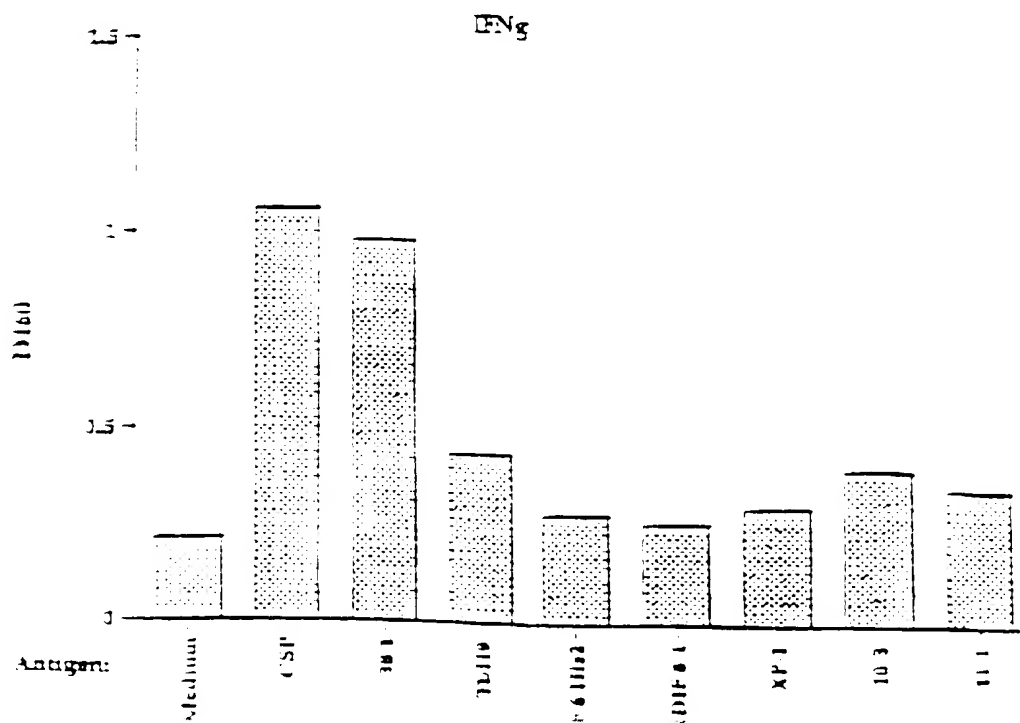
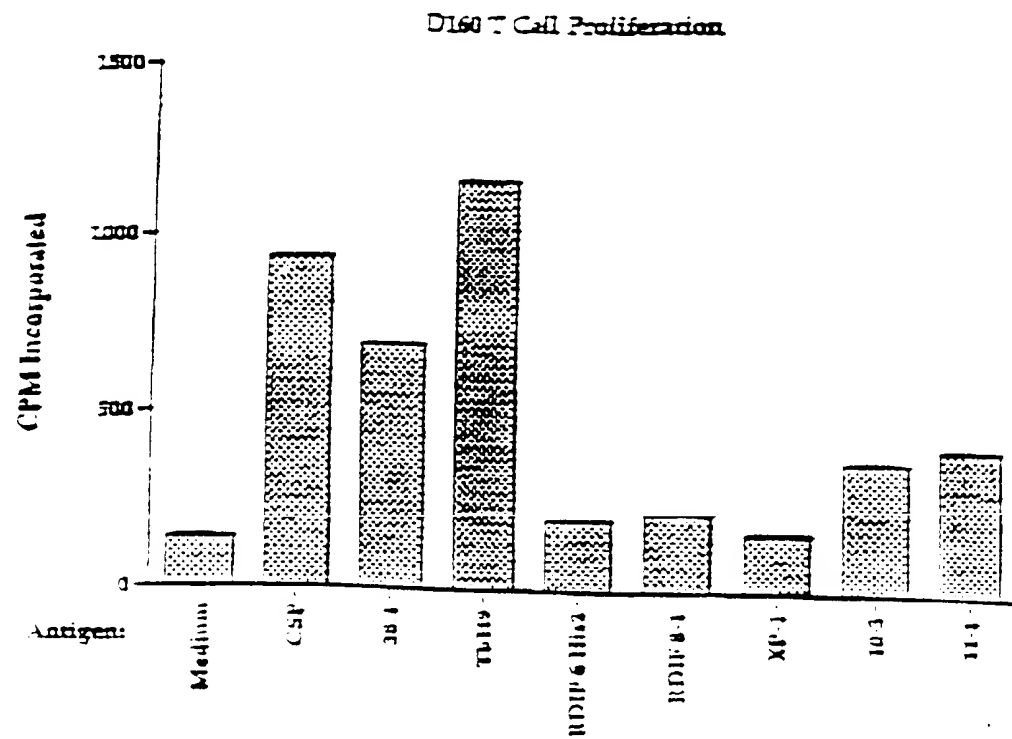


D184 IFN γ

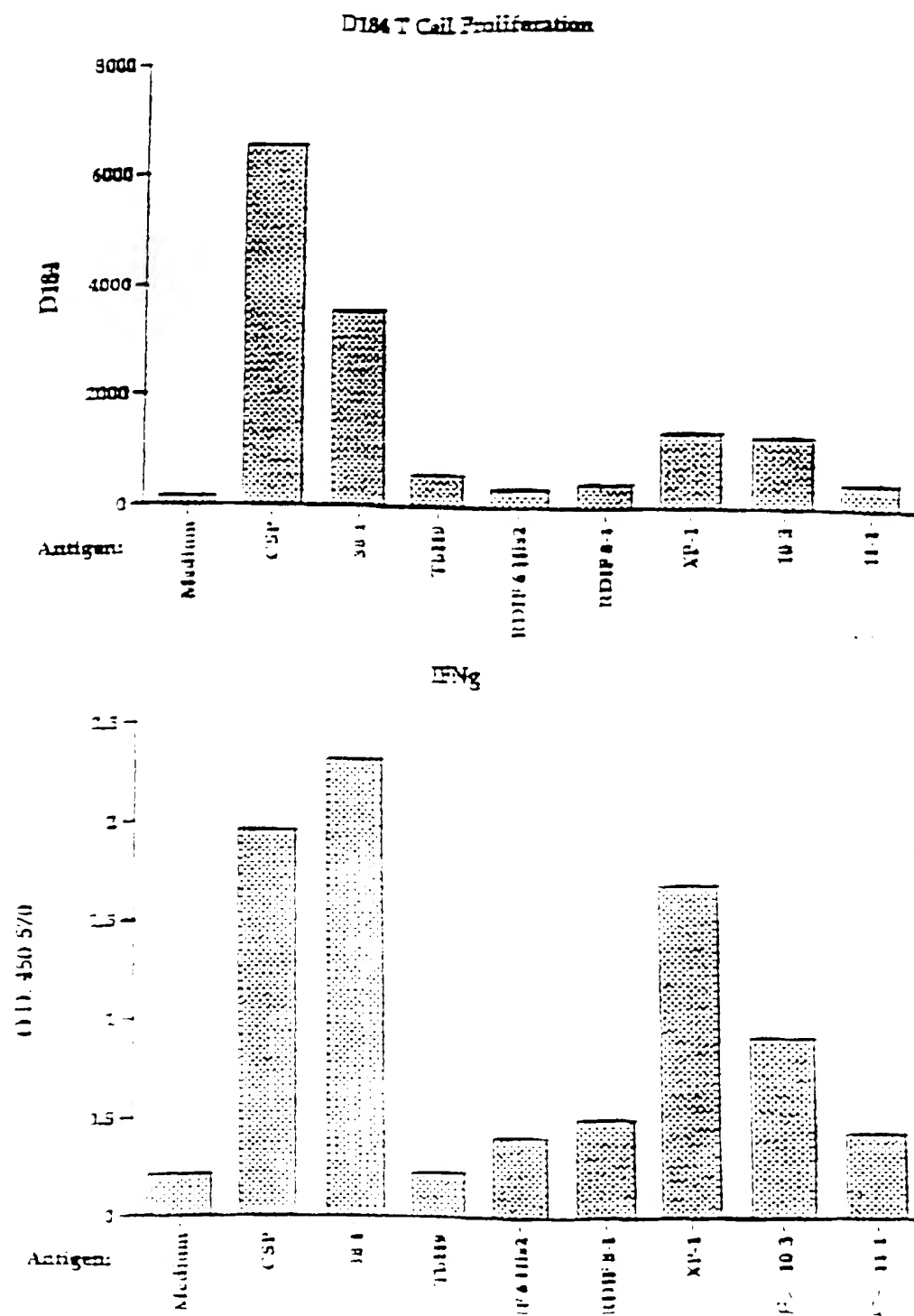




FIGS. 7A-B

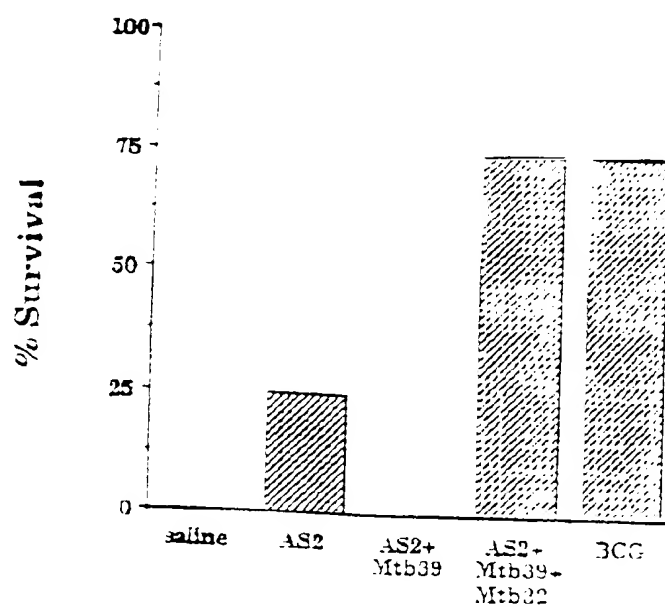


FIGS 8A-B

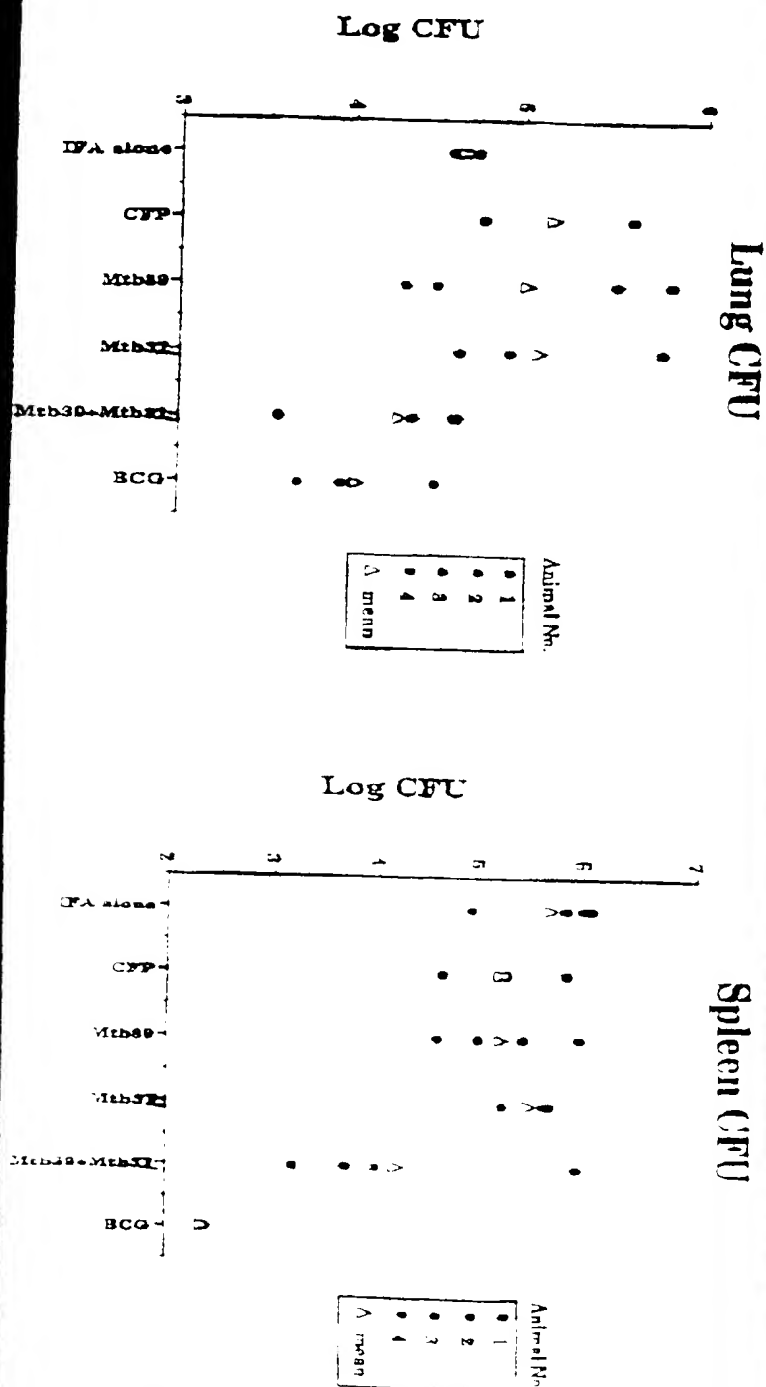


FIGS. 9A-B

**Tuberculosis: Protection of Cynomolgus Monkeys
with Recombinant Antigens of Mtb**



Aerosol TTB Challenge of Vaccinated Guinea Pigs



FIGS. 11A-B

DNA Immunized mice challenged with aerosol TB (lung CFU)

